Philipp Strazny

Editor

encyclopedia of
LINGUISTICS

Volume 1

A-L

FITZROY DEARBORN

An Imprint of the Taylor & Francis Group
New York • Oxon

encyclopedia of

LINGUISTICS

Volume 1

Philipp Strazny

Editor

encyclopedia of

LINGUISTICS

Volume 1

A-L

FITZROY DEARBORN

An Imprint of the Taylor & Francis Group
New York • Oxon

Board of Advisers

Contents

Alphabetical List of Entries

Alphabetical List of Entries

Alphabetical List of Entries

Alphabetical List of Entries

Thematic List of Entries

Thematic List of Entries

Thematic List of Entries

Thematic List of Entries

Color Terms
Compositional Semantics
Context
Coordination
Definiteness
Figurative Speech
Genericity
Idioms
Kinship Terms
Meaning
Metaphor
Metonymy
Modification
Naming
Onomatopoeia
Plurality
Proper Nouns
Quantification
Reference
Register
Sapir–Whorf Hypothesis
Sign Relationship
Time and Tense

SIGN LANGUAGES
American Sign Language
British Sign Language
Indo-Pakistani Sign Language
Japanese Sign Language
Signed Languages

SINO-TIBETAN LANGUAGES
Burmese
Chinese: Mandarin
Tibetan

SOCIOLINGUISTICS
African American Vernacular English
Courtroom Discourse
Diglossia
Emotion and Language
Ethnicity and Language
Ethnography of Communication
Euphemism
Gender and Language
Identity and Language
Ideology and Language
Mass Media and Language
Medicine and Language

Personality and Language
Sociolect
Standard Language
Variation

SYNTAX
Anaphora
Aspect
Auxiliaries
Case
Clause
Configurationality
Constituency Test
Deep Structure
Determiner
Empty Categories
Focality
Function Words
Grammatical Function
Iconicity
Long-Distance Dependency
Mood
Noun Incorporation
Phrase Structure
Possessives
Predication
Pro-forms
Questions
Serial Verb Constructions
Story Grammar
Syntactic Categories
Tense: Syntax
Thematic Structure
Universal Grammar
Voice
Word Order

TAI LANGUAGES
Thai and Tai Languages

TYPOLOGY
Morphological Typology
Semantic and Discourse Typology
Syntactic Typology

URALIC LANGUAGES
Finnish and Finnic Languages
Hungarian and Ugric Languages
Udmurt

Contributors

Abazova, Alfia. Jersey City, New Jersey.

Ackema, Peter. University of Edinburgh, UK.

Aikhenvald, Alexandra. Research Center for Linguistic Typology, Institute for Advanced Study, La Trobe University, Melbourne, Australia.

Altarriba, Jeanette. Department of Psychology, University at Albany, State University of New York, USA.

Altmann, Lori J.P. Department of Communication Sciences and Disorders, University of Florida, USA.

Anderson, Warren D. Department of Foreign Languages and Anthropology, Southeast Missouri State University, USA.

Antieau, Lamont. Linguistics Program, University of Georgia, USA.

Archibald, John. Department of Linguistics, University of Calgary, Alberta, Canada.

Arends, Jacques. Department of Linguistics, University of Amsterdam, The Netherlands.

Aronoff, Mark. Department of Linguistics, Stony Brook University, Stony Brook, New York, USA.

Attardo, Salvatore. English Department, Youngstown State University, Ohio.

Axelrod, Melissa. Department of Linguistics, University of New Mexico, USA.

Baldauf, Richard B., Jr. School of Education, University of Queensland, Australia.

Baptista, Marlyse. Linguistics Program, University of Georgia, USA.

Barcelona, Antonio. Departamento de Filología Inglesa, Universidad de Murcia, Spain.

Barnard, Roger. Department of General and Applied Linguistics, University of Waikato, New Zealand.

Barron, Anne. Englisches Seminar, Universität Bonn, Germany.

Barry, Betsy. University of Georgia, USA.

Barss, Andrew. Department of Linguistics, University of Arizona, USA.

Barton, Ellen. Linguistics Program, Department of English, Wayne State University, Michigan.

Bashir, Elena. University of Chicago, Illinois.

Bastiaanse, Roelien. University of Groningen, The Netherlands.

Basturkmen, Helen. Department of Applied Language Studies and Linguistics, Auckland University, New Zealand.

Bat-El, Outi. Department of Linguistics, Tel-Aviv University, Israel.

Battaner-Moro, Elena. Departamento de Lengua Española y Lingüística General, Universidad Nacional de Educación a Distancia, Madrid, Spain.

Bauer, Laurie. Linguistics and Applied Language Studies, Victoria University of Wellington, New Zealand.

Baumgartner, Joanne Marie. Australian National University, Canberra, Australia.

Contributors

Beckwith, Christopher I. Indiana University, USA.

Belnap, R. Kirk. Brigham Young University, Utah.

Benner, Allison. Victoria, British Columbia, Canada.

Benremouga, Karima. Fine Arts and Language Arts, San Jacinto College Central, Pasadena, TX.

Berge, Anna. Alaska Native Language Center, University of Alaska, Fairbanks, USA.

Bernhardt, Karl. University of Buckingham, United Kingdom.

Bhatia, Tej K. Syracuse University, New York.

Bird, Steven. Linguistic Data Consortium, University of Pennsylvania, Philadelphia.

Blume, Kerstin. Bergische Universität Wuppertal, Germany.

Bodomo, Adams. Department of Linguistics, University of Hong Kong, China.

Bolotin, Naomi. Cambridge, Massachusetts.

Born, Renate. Department of Germanic and Slavic Languages, University of Georgia, USA.

Bowden, John. Linguistics Department, Research School of Pacific and Asian Studies, The Australian National University, Australia.

Bretones, Carmen. International Computer Science Institute, Berkeley, California.

Brown, Jeffrey M. Texas A&M International University, USA.

Branigan, Holly. School of Philosophy, Psychology, and Language Sciences, University of Edinburgh.

Burley, Lynn Ann. Department of Writing and Speech, University of Central Arkansas, USA.

Camdzic, Amela. University College London, United Kingdom.

Campana, Mark. Kobe, Hyogo-ken, Japan.

Carnie, Andrew. Department of Linguistics, University of Arizona, USA.

Carstairs-McCarthy, Andrew. Linguistics Department, University of Canterbury, Christchurch, New Zealand.

Carter, Michael G. Department of Oriental and East European Studies, Oslo University, Norway.

Clankie, Shawn M. Center for Language Studies, Otaru University of Commerce, Japan.

Cloutier, Robert A. Amsterdam Center for Language and Communication/Department of Dutch Studies, University of Amsterdam, Amsterdam, The Netherlands.

Cockerton, Camilla M. Happy Valley, Rakaia, New Zealand.

Cogill-Koez D. Department of Linguistics, University of New England, Armidale, Australia.

Cohen, Ariel. Department of Foreign Literatures and Linguistics, Ben-Gurion University of the Negev, Beer Sheva, Israel.

Colarusso, John. McMaster University, Hamilton, Ontario, Canada.

Cole, Jennifer. Department of Linguistics, University of Illinois at Urbana-Champaign, USA.

Colley, Michael. Linguistics Program, University of Georgia, USA.

Comrie, Bernard. Max-Planck-Institut für evolutionäre Anthropologie, Leipzig, Germany.

Corré, Alan D. Hebrew Studies, University of Wisconsin-Milwaukee, USA.

Cots, Josep Maria. Departament d'Anglès i Lingüística, Universitat de Lleida, Catalunya, Spain.

Croft, William. Department of Linguistics, University of Manchester, United Kingdom.

Crowley, Terry. Department of General and Applied Linguistics, University of Waikato, New Zealand.

Daniels, Peter T. Bronx, New York.

Daniliuc, Radu. Department of Linguistics, Australian National University, Canberra, Australia.

Daniliuc, Laura. Department of Linguistics, Australian National University, Kambah, Australia.

Daza, Jaime Luis. Linguistics Department, Unversity of California, Los Angeles, USA.

Dimitriu, Rodica. Department of English, University of Iasi, Romania.

Ding, Picus Sizhi. Macao, Hong Kong, China.

Dooley Collberg, Sheila. Department of Linguistics, University of Arizona, USA.

Douglas, Dan. Department of English, Iowa State University, USA.

Dumas, Bethany K. Department of English, University of Tennessee, USA.

Ebermann, Erwin. Department of African Studies, University of Vienna, Austria.

Echeruo, Michael J.C. Department of English, Syracuse University, Syracuse, New York, USA.

Eira, Christina. Adelaide Graduate Centre, University of Adelaide, Australia.

Engelberg, Stefan. University of Wuppertal, Germany.

Ennaji, Moha. Fez, Morocco.

Essegbey, James. Department of African Linguistics, Leiden University, The Netherlands.

Evans, Nicholas. Department of Linguistics & Applied Linguistics, University of Melbourne, Australia.

Falk, Julia S. La Jolla, California, USA.

Feldman, Laurie Beth. University of Albany, State University of New York, USA.

Fettes, Mark. Simon Fraser University, Ottawa, Ontario, Canada.

Flier, Michael S. Department of Slavic Languages and Literatures, Harvard University, Massachusetts, USA.

Foster-Cohen, Susan H. The Champion Centre and University of Canterbury, Christchurch, New Zealand.

Gagné, Christina L. Department of Psychology, University of Alberta, Edmonton, Canada.

Gildea, Spike. Department of Linguistics, University of Oregon, USA.

Goatly, Andrew. Department of English, Lingnan University, N.T., Hong Kong, China.

González-Romero, Luisa. Departamento de Filología Inglesa, Universidad de Huelva, Andalucía, Spain.

Gordon, Lynn. Department of English, Washington State University, USA.

Graesser, Arthur. Department of Psychology, University of Memphis, USA.

Green, Ian. Graduate School, University of Tasmania, Hobart, Australia.

Greenberg, Marc L. Slavic Department, University of Kansas, USA.

Grenoble, Lenore A. Program in Linguistics and Cognitive Science, Dartmouth College, New Hampshire.

Gribble, Charles. Slavic Department, Ohio State University, USA.

Grohmann, Kleanthes K. Department of English Studies, University of Cyprus, Nicosia.

Grondona, Veronica M. Department of English, Eastern Michigan University, USA.

Guillaume, Jacques. Paris, France.

Gutiérrez-Rexach, Javier. Department of Spanish & Portuguese, Ohio State University, USA.

Haberlandt, Karl. Department of Psychology, Trinity College, Connecticut, USA.

Hahn, Udo. Linguistische Informatik/Computerlinguistik, Freiburg University, Germany.

Hajicova, Eva. MFF UK, Prague, Czech Republic.

Hall, David R. Linguistics Department, Macquarie University Sydney, Australia.

Hall, Kira. Department of Linguistics, Stanford University, California.

Hammond, Michael. Department of Linguistics, University of Arizona, USA.

Harlow, Ray. Department of General and Applied Linguistics, University of Waikato, New Zealand.

Harrington, Jonathan. Institute of Phonetics and Digital Speech Processing, University of Kiel, Germany.

Contributors

Harrison, K. David. Linguistics Department, Swarthmore College, Swarthmore, Pennsylvania, USA.

Hartig, Matthias. Sprach- und Literaturwissenschaft, Universität Paderborn, Germany.

Hartsuiker, Robert J. Department of Experimental Psychology, Ghent University, Belgium.

Heine, Bernd. Institut für Afrikanistik, Universität zu Köln, Germany.

Hendriks, Petra. Center for Language and Cognition Groningen, University of Groningen, The Netherlands.

Heredia, Roberto R. Texas A&M International University, USA.

Herrmann, Stefanie. University of Tübingen, Baden-Württemberg, Germany.

Holtgraves, Thomas M. Psychological Sciences, Ball State University, Indiana.

Hongladarom, Krisadawan. Department of Linguistics, Faculty of Arts, Chulalongkorn University, Bangkok, Thailand.

Hopple, Paulette. Summer Institute of Linguistics, University of Texas, Arlington, USA.

Howe, Darin. Department of Linguistics, University of Calgary, Alberta, Canada.

Hua, Zhu. Department of Speech, University of Newcastle upon Tyne, United Kingdom.

Hudson, Grover. Linguistics Department, Michigan State University, USA.

Hurford, James R. Department of Linguistics, University of Edinburgh, Scotland, United Kingdom.

Hutchins, W. John. Norwich, United Kingdom.

Ibarretxe-Antunano, Iraide. Facultad de Filosofía y Letras, Universidad de Deusto - Deustuko Unibertsitatea, Bizkaia, Spain.

Inchaurralde, Carlos. Departamento Filologia Inglesa y Alemana, Universidad de Zaragoza, Spain.

Insler, Stanley. Department of Linguistics, Yale University, Connecticut.

Jaggar, Philip J. School of Oriental & African Studies, London, United Kingdom.

Jensen, Cheryl. Summer Institute of Linguistics, Belém, Pará, Brazil.

Johnson, E. Marcia. Department of General & Applied Linguistics, University of Waikato, New Zealand.

Johnson, Ellen. Department of English, Rhetoric, and Writing, Berry College, Georgia.

Jones, Francis R. School of Modern Languages, University of Newcastle, United Kingdom.

Jongman, Allard. Department of Linguistics, University of Kansas, USA.

Justus, Carol F. Classics Department, University of Texas, Austin, USA.

Karimi, Simin. Department of Linguistics, University of Arizona, USA.

Katz, Leonard. Haskins Laboratories, New Haven, Connecticut.

Kaye, Alan S. Linguistics Department, California State University-Fullerton, USA.

Kemper, Susan. Department of Psychology, University of Kansas, USA.

Kibrik, Andrej. Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

King, Robert D. Department of Linguistics, University of Texas, Austin, USA.

Kirtchuk-Halevi, Pablo Isaac. Beer Sheva, Israel.

Knight, Chris. School of Social Sciences, University of East London, Dagenham, Essex, United Kingdom.

Kroskrity, Paul V. Department of Anthropology, University of California, Los Angeles, USA.

Kull, Kalevi. Department of Semiotics, University of Tartu, Estonia.

Kullavanijaya, Pranee. Department of Linguistics, Faculty of Arts, Bangkok, Thailand.

Labeau, Emmanuelle. School of Languages and Social Sciences, Aston University, Birmingham, West Midlands, United Kingdom.

Lahousse, Karen. Katholieke Universiteit Leuven (Catholic University Leuven) & Fonds voor, Wetenschappelijk Onderzoek–Vlaanderen (Fund of Scientific Research–Flanders).

Lainio, Jarmo. Centre for Finnish Studies, Mälardalen University, Eskilstuna/Västerås, Swedan.

Lang, George. Dean of Arts, University of Ottawa, Canada.

Lasnik, Howard. University of Maryland, USA.

Leiber, Justin. Philosophy Department, University of Houston, Texas.

Leitner, Gerhard. Institut fuer englische Philologie, Freie Universität Berlin, Germany.

Levis, John M. Department of English, Iowa State University, USA.

Lieberman, Philip. Department of Cognitive and Linguistic Sciences, Brown University, Rhode Island.

Llurda, Enric. Departament d'Anglès i Linguistica, Universitat de Lleida, Catalunya, Spain.

Loengarov, Alexander. Research Assistant of the Fund for Scientific Research–Flanders (Belgium). Department of Linguistics, K.U. Leuven, Belgium.

Lohr, Marisa. Trinity College, Cambridge, United Kingdom.

Louwerse, Max M. Department of Psychology/Institute for Intelligent Systems, University of Memphis, USA.

Luo, Yongxian. Melbourne Institute of Asian Languages and Societies, University of Melbourne, Australia.

Luraghi, Silvia. Dipartimento di Linguistica, Università di Pavia, Italy.

Mackert, Michael. German-English Language Services, Morgantown, West Virginia, USA.

MacMahon, Michael K.C. Department of English Language, University of Glasgow, Scotland, United Kingdom.

Mahanta, Deepshikha. Central Institute of English and Foreign Languages, Hyderabad, India.

Mahdi, Waruno. Berlin, Germany.

Mahmoodi-Bakhtiari, Behrooz. Center for the Great Islamic Encyclopedia, Tehran, Iran.

Marret, Carine. Nice, France.

Martín Arista, Javier. Departamento de Filologías Modernas, Universidad de La Rioja, Logroño, Spain.

Martín Camacho, José Carlos. Facultad de Filosofía y Letras, Universidad de Extremadura, Cáceres, Spain.

Mattina, Anthony. Department of Anthropology, University of Montana, USA.

Mattissen, Johanna. Institut für Linguistik (Department of Linguistics), University of Köln (Cologne), Germany.

Maxwell, Judith M. Department of Anthropology, Tulane University, Louisiana.

Mc Laughlin, Fiona. African Languages and Linguistics, University of Florida, USA.

McDonald, David. Arlington, Massachusetts.

McGregor, William B. Department of Linguistics, University of Aarhus, Denmark.

McMenamin, Gerald. Fresno, California.

Meisel, Jürgen M. Institut für Romanistik, Universität Hamburg, Germany.

Melançon, Megan E. Department of English, Speech, and Journalism, Georgia College and State University, USA.

Menchetti, Angiolo. Dipartimento di Scienze Storiche del Mondo Antico, Università di Pisa, Italy.

Miller, Catherine. Iremam-MMSH, University of Aix en Provence, France.

Minkova, Donka. Department of English, University of California, Los Angeles, USA.

Mithun, Marianne. Department of Linguistics, University of California, Santa Barbara, USA.

Moore, John. Department of Linguistics, University of California, San Diego, USA.

Morgan, Michael. Kobe City University of Foreign Studies, Kobe, Japan.

Mufwene, Salikoko. Department of Linguistics, University of Chicago, Illinois, USA.

Contributors

Mukherjee, Joybrato. Department of English, Justus Liebig University, Giessen, Germany.

Müller, Horst M. AG Experimentelle Neurolinguistik, Fakultät für Linguistik und Literaturwissenschaft, Universität Bielefeld, Germany.

Murray, Robert W. Department of Linguistics, University of Calgary, Alberta, Canada.

Murray, Denise. National Centre for English Language Teaching and Research, Macquarie University, Ryde, Australia.

Nash, David. Australian National University, Canberra, Australia.

Nau, Nicole. General Linguistics, Kiel University, Germany.

Newmark, Leonard. University of California, San Diego, USA.

Nichols, Lynn. Department of Linguistics, University of California, Berkeley.

Nifadopoulos, Christos G. Bristol, United Kingdom.

Noma, Hideki. Department of Korean Studies, Tokyo University of Foreign Studies, Japan.

O'Rourke, Sean. Department of Linguistics, Yale University, Connecticut.

Oakes, Michael P. University of Sunderland, United Kingdom.

Obeng, Samuel Gyasi. Linguistics Department, Indiana University, USA.

Ohala, John J. Department of Linguistics, University of California at Berkeley, USA.

Oliverio, Giulia. Alaska Native Language Center, University of Alaska-Fairbanks, USA.

Ongstad, Sigmund. Oslo University College, Norway.

Ọla Orie, Ọlanikẹ. Linguistic Program & Department of Anthropology, Tulane University, Louisiana.

Orr, Robert. Gloucester, Ontario, Canada.

Otero, Carlos. University of California, Los Angeles, USA.

Oventile, Robert S. Pasadena City College, California.

Paap, Kenneth R. Department of Psychology, New Mexico State University, USA.

Paltridge, Brian. OE Faculty of Education, University of Sydney.

Pandey, Anjali. Department of English, Salisbury University, Maryland, USA.

Pandey, Anita. Department of English & Language Arts, Morgan State University, Maryland

Papangeli, Dimitra. Faculteit der Letteren, Utrecht Institute of Linguistics, OTS, The Netherlands.

Papen, Robert A. Linguistique et de didactique des langues, Université du Québec, Montréal, Canada.

Parasher, Shree Vallabh. Central Institute of English and Foreign Languages, Hyderabad, India.

Parkvall, Mikael. Institutionen för Lingvistik, Stockholms Universitet, Sweden.

Pensalfini, Rob. School of English, Media Studies and Art History, University of Queensland, Brisbane, Australia.

Pepperberg, Irene M. Department of Psychology, Brandeis University, Waltham, MA 02454.

Perkins, Michael. Human Communication Sciences, University of Sheffield, South Yorkshire, United Kingdom.

Pierce, Marc. Department of Germanic Languages and Literatures, University of Michigan, USA.

Piller, Ingrid. Linguistics Department, University of Sydney, Australia.

Pineda, Baron. Department of Anthropology, Oberlin College, Ohio.

Pires, Acrisio. Department of Linguistics, University of Michigan, USA.

Ponchon, Thierry. Université de Reims, France.

Prakasam, Vennelakanti. Central Institute of English and Foreign Languages, Lucknow, India.

Quesada, J. Diego. Escuela de Literatura y Ciencias del Lenguaje. Universidad Nacional. Heredia. Costa Rica.

Radev, Dragomir R. School of Information, Department of Electrical Engineering and Computer Science, and Department of Linguistics, University of Michigan, USA.

Rajagopalan, Kanavillil. State University at Campinas, Campinas, Brazil.

Rathert, Monika. Seminar für Sprachwissenschaft, Universität Tübingen, Baden-Württemberg, Germany.

Reh, Mechthild. Department of African Languages and Ethiopian Studies, Hamburg University, Germany.

Reid, Nicholas. School of Languages, Cultures and Linguistics, University of New England, Armidale, Australia.

Reid, Lawrence A. Tokyo, Japan.

Reintges, Chris H. Center for Linguistics, Leiden University, The Netherlands.

Reiss, Charles. Linguistics Program, Concordia University, Montreal, Canada.

Riedinger, Edward A. Department of Spanish and Portugese, The Ohio State University, USA.

Rodríguez-Arrizabalaga, Beatriz. Departamento de Filología Inglesa. University of Huelva, Andalucía, Spain.

Rogers, Henry. Department of Linguistics, University of Toronto, Ontario, Canada.

Romaine, Suzanne. Merton College, University of Oxford, United Kingdom.

Rubio, Gonzalo. Department of Classics and Ancient Mediterranean Studies, Pennsylvania State University, USA.

Rutkowski, Paweł. Department of Polish, Warsaw University, Poland.

Sackmann, Robin. Freie Universität Berlin, FB PhilGeist, Freie Universität Berlin, Germany.

Sanders, Alton. Department of Computer Science and Systems Analysis, Miami University, Ohio.

Sanders, Carol. Department of Linguistic, Cultural and International Studies, University of Surrey, Guildford, United Kingdom.

Sanders, Ruth H. Department of German, Russian, East Asian and Hebrew, Miami University, Ohio.

Sansò, Andrea. Dipartimento di Linguistica, Universita' di Pisa, Italy.

Savova, Lilia. Department of English, Indiana University of Pennsylvania, USA.

Sayahi, Lotfi. Department of Languages, Literatures and Cultures, The University at Albany, State University of New York, USA.

Schneider, Klaus P. Englisches Seminar, Universität Bonn, Germany.

Scott, Suzanne. Dunedin, New Zealand.

Sew, Jyh Wee. Singapore.

Sezer, Engin. Bilkent University, Ankara, Turkey.

Sgall, Prague Petr. Faculty of Mathematics and Physics, Prague, Czech Republic.

Sgarbas, Kyriakos N. Department of Electrical and Computer Engineering, University of Patras, Greece.

Shekar, Chandra. California State University, Fresno.

Sicoli, Mark. Departments of Anthropology and Linguistics, University of Michigan, USA.

Siegel, Jeff. School of Languages, Cultures and Linguistics, University of New England, Australia, and Department of Second Language Studies, University of Hawai'i, USA.

Signorini, Inês. Department of Applied Linguistics, Universidade Estadual de Campinas (UNICAMP), São Paulo, Brazil.

Silzer, Peter J. Department of TESOL and Applied Linguistics, Biola University, USA.

Siple, Patricia. Department of Psychology, Wayne State University, Michigan.

Sitzmann, Alexander. Institut für Germanistik, Universität Wien, Austria.

Sivell, John N. Department of Applied Language Studies, Brock University, St. Catharines, Ontario, Canada.

Smith, Ian R. Department of Languages, Literatures and Linguistics, York University, Toronto, Ontario, Canada.

Smith, Allison. Department of English, Louisiana Tech. University, USA.

Snoj, Marko. Institut za slovenski jezik ZRC SAZU, Slovenia.

Sotillo, Susana M. Department of Linguistics, Montclair State University, New Jersey.

Contributors

Steinkrüger, Patrick O. Zentrum für allgemeine Sprachwissenschaft, Berlin, Germany.

Strazny, Philipp. Monona, USA.

Sujoldzic, Anita. Institute for Anthropological Research, Zagreb, Croatia.

Sunderland, Jane. Department of Linguistics and English Language, Bowland College, Lancaster University, Lancashire, United Kingdom.

Tadmor, Uri. Department of Linguistics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany/Jakarta, Indonesia.

Tambovtsev, Yuri. Department of English and Linguistics, KFNPU, Novosibirsk, Russia.

Tatham, Mark. Colchester, United Kingdom.

ten Hacken, Pius. School of European Languages, University of Wales Swansea, UK.

Thomason, Sarah G. University of Michigan, USA.

Thomassen, Arnold J.W.M. University of Nijmegen, The Netherlands.

Toribio, Josefa. Department of Philosophy, The University of Edinburgh, Scotland.

TSE, Grace Yuen Wah. School of Arts and Social Sciences, The Open University of Hong Kong, China.

Tsedryk, Egor. Department of French Studies, University of Western Ontario, London, Canada and Department of French Studies, Concordia University, Montreal, Canada.

Tserdanelis, Giorgos. The Ohio State University, USA.

Turovski, Aleksei. Tallinn Zoo, Tallinn, Estonia.

van Gelderen, Elly. English Department, Arizona State University, USA.

van Wieringen, Astrid. Lab. Exp. ORL, Katholieke Universiteit Leuven, Belgium.

Vaux, Bert. Department of Foreign Languages and Linguistics, University of Wisconsin-Milwaukee, USA.

Veselinovic, Elvira M. University of Cologne, Germany.

Viereck, Wolfgang. University of Bamberg, Germany.

Voeltz, F.K. Erhard. Institut für Afrikanistik, Universität zu Köln, Germany.

Vovin, Alexander. Department of East Asian Languages and Literatures, University of Hawai'i, USA.

Vrooman, Michael. Department of Modern Languages and Literatures, Grand Valley State University, Michigan.

Wagner, Richard. Department of Psychology, Florida State University, USA.

Walker, Douglas C. Department of French, Italian and Spanish, University of Calgary, Alberta, Canada.

Walker, James A. Department of Languages, Literatures, and Linguistics, York University, Toronto, Ontario, Canada.

Wang, William S.-Y. Department of Electronic Engineering, City University of Hong Kong, China.

Watt, William C. Social Science, University of California, Irvine, USA.

Waugh, Linda R. Department of French, Italian and English, University of Arizona, USA.

Wehmeyer, Ann. Department of African and Asian Languages and Literatures, University of Florida, USA.

Weiss, Sabine. Brain Research Institute, Cognitive Neuroscience Group, Medical University of Vienna, Austria.

Wexler, Paul. Department of Linguistics, Tel-Aviv University, Israel.

Whaley, Lindsay. Linguistics & Cognitive Science, Dartmouth College, New Hampshire.

Wilcox, Phyllis Perrin. Department of Linguistics, University of New Mexico, USA.

Wilson, Andrew. Linguistics Department, Lancaster University, United Kingdom.

Wireback, Kenneth J. Department of Spanish and Portuguese, Miami University, Ohio.

Woll, Bencie. Sign Language & Deaf Studies, Department of Language and Communication Science, City University, London, United Kingdom.

Wong, Andrew. Department of Linguistics, University of Hawai'i at Manoa, USA.

Woodman, Karen. School of Languages, Cultures and Linguistics, University of New England, Armidale, Australia.

Zeshan, Ulrike. Research Centre for Linguistic Typology, La Trobe University, Bundoora, Australia.

Introduction

The study of language goes far back in recorded history. Almost two-and-a-half millenia ago, the Indian grammarian Panini wrote his formal treatises on Sanskrit, Xun Zi appeared as China's first major philosopher of language, and Plato and Aristotle initiated the Greek philosophy of language. Since the renaissance, there has been an increasing focus on the description of individual languages, the exploration of familial relationships between languages, and the formulation of increasingly general theories of language structure. Now, almost every university or college has a language department or even a specialized linguistics department, which means that an immense number of researchers are working in the field and have published an enormous body of primary literature.

Given this wealth of published data, it is not surprising to meet beginning graduate students of linguistics who already identify themselves as 'syntacticians' or 'phoneticians'. This early specialization reflects genuine interest, but is also in part a mechanism by which students block out a large number of possible inputs to be able to concentrate on a more manageable few. In their subsequent research, most researchers narrow down their field of interest even more; they become immersed in the highly conceptual and terminological world of their specialty, and they often write articles comprehensible only to their small group of peers. In other words, linguistics is a mature science and, as in other scientific fields, there can be a communications gap both within the field and, even more so, between the active researchers and the general public.

The *Encyclopedia of Linguistics* provides an accessible overview of and introduction to the multiple facets of the study of language. To bridge this gap between professional linguists and the general public, my editorial colleagues and I made this encyclopedia very readable by eliminating technical terminology as far as possible and by making each essay self-contained.

How to Use This Book

The *Encyclopedia of Linguistics* is organized into a series of 508 free-standing essays, between 1000 and 3000 words in length. They range from factual narrative entries to thematic and analytical discussions, and combinations of all these. Where debates and controversies occur, these are indicated and discussed. As far as possible, this book takes the field of linguistics up to the present, at least to the opening years of the twenty-first century.

Perhaps the most significant feature of the encyclopedia is the easily accessible **A to Z format**. Cross-referencing in the form of **See Also**s at the end of each entry refers the reader to other related essays. A thorough, analytical **Index** complements the accessibility of the entries, easing the reader's entry into the wealth of information provided. **References** at the end of each entry refer the reader to seminal writings as well as some of the most recent work on the subject. Other special features include **12 language-distribution maps** and a **thematic Table of Contents** in addition to an **alphabetical Table of Contents.** In addition, **more than 100 illustrations** are dispersed throughout. A total of **288 scholars** from **34 countries** have contributed their expertise to this encyclopedia.

Contents

These essays have been written by scholars who provide a general introduction to the material without presupposing knowledge about the subject and without going into a

theoretical depth that would raise questions that cannot be answered in the space given. The entries fall into the following thematic categories: linguistic topics (for example, code switching) (50%), languages (for example, Sumerian) (30%), persons (for example, Noam Chomsky) (15%), and regions of the world (for example, Algeria) (5%).

Linguistic Topics

Among these essays are general introductions to major fields of inquiry, such as semantics, historical linguistics, and neurolinguistics. Other articles concentrate on issues within those fields, introducing concepts that are important in linguistics regardless of theoretical perspective, such as 'affix', or 'reference', or which are so commonly used that the concept becomes unquestioned, such as 'phoneme' or 'deep structure'. Other articles describe crucial phenomena that any theory of language has to explain, such as 'acquisition', 'tone', or 'aphasia'.

Languages

These essays cover the major language families of the world and discuss how individual languages are related to each other. Articles concentrating on specific languages from Ainu to Zuni explain where these languages are spoken, by whom, and under what sociopolitical circumstances. They provide a glimpse of the language's structure and highlight particularly interesting characteristics with examples.

Persons

These essays highlight the major theoretical contributions of noted linguists, discuss the influences that led up to their work, and put the contributions into historical perspective, in addition to providing basic biographical sketches.

Regions

These essays map out the linguistic landscape of a particular region. They show which languages are spoken, explain how this particular situation came about historically, and discuss language-political issues relevant in this region.

We have made a concerted effort to cover languages around the globe. Should you note any imbalance in favor of 'western' regions, languages, persons, or topics, this simply reflects that the European languages are the best-studied languages in the world.

Acknowledgments

A

Acoustic Phonetics

Acoustic phonetics is the study of the acoustic characteristics of speech. Speech consists of variations in air pressure that result from physical disturbances of air molecules caused by the flow of air out of the lungs. This airflow makes the air molecules alternately crowd together and move apart (oscillate), creating increases and decreases, respectively, in air pressure. The resulting sound wave transmits these changes in pressure from speaker to hearer. Sound waves can be described in terms of physical properties such as cycle, period, frequency, and amplitude. These concepts are most easily illustrated when considering a simple wave corresponding to a pure tone. A cycle is a sequence of one increase and one decrease in air pressure. A period is the amount of time (expressed in seconds or milliseconds) that one cycle takes. Frequency is the number of cycles in one second, expressed in hertz (Hz). An increase in frequency usually results in an increase in perceived pitch. Amplitude refers to the magnitude of vibrations, with larger vibrations resulting in greater peaks of pressure (greater amplitude), which usually result in an increase in perceived loudness.

Unlike pure tones, which rarely occur in the environment, speech sounds are complex waves with combinations of different frequencies and amplitudes. However, as first stated by the French mathematician Fourier (1768–1830), any complex wave can be described as a combination of simple waves. A complex wave has a regular rate of repetition, known as the fundamental frequency ($F0$). Changes in $F0$ give rise to differences in perceived pitch, whereas changes in the number of constituent simple waves and their amplitude relations result in perceived differences in timbre or quality.

Fourier's theorem enables us to describe speech sounds in terms of the frequency and amplitude of each of its constituent simple waves. Such a description is known as the spectrum of a sound. A spectrum is visually displayed as a plot of frequency vs. amplitude, with frequency represented from low to high along the horizontal axis and amplitude from low to high along the vertical axis.

The usual energy source for speech is the airstream generated by the lungs. This steady flow of air is converted into brief puffs of air by the vibrating vocal folds, two muscular folds housed in the larynx. The dominant way of conceptualizing the process of speech production is in terms of the source-filter theory, according to which the acoustic characteristics of speech can be understood as a result of a source component and a filter component. The source component is determined by the rate of vocal fold vibration, which in turn is affected by a number of factors, including the rate of airflow and the mass and stiffness of the vocal folds. The rate of vocal fold vibration directly determines the $F0$ of the waveform. The mean $F0$ for adult women is approximately 220 Hz, and approximately 130 Hz for adult men. In addition to their role as properties of individual speech sounds, $F0$ and amplitude also signal emphasis, stress, and intonation.

For speech, the source component itself has a complex waveform, and its spectrum will typically show the highest energy at the lowest frequencies and a number of higher frequency components that systematically decrease in amplitude. This source component is subsequently modified by the vocal tract above the larynx, which acts as the filter. This filter
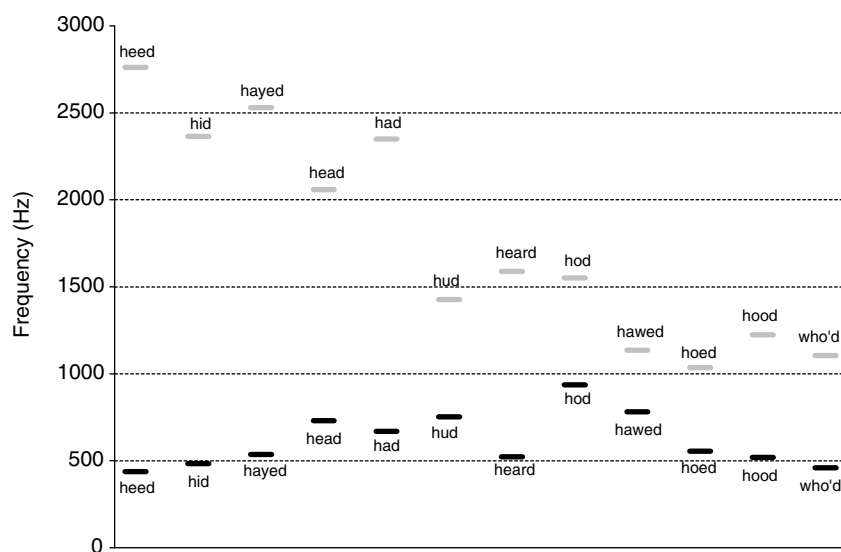
Figure 1. Frequencies of the first two formants of 12 vowels of American English, averaged across 48 adult female speakers. *F*1 is in black, *F*2 is in gray.
(Source: Hillenbrand et al., J. Acoust. Soc. Am., 1995)

enhances energy in certain frequency regions and suppresses energy in others, resulting in a spectrum with peaks and valleys, respectively. The peaks in the spectrum (local energy maxima) are known as formant frequencies. The lowest-frequency peak is known as the first formant, or $F1$, the next lowest is $F2$, and so on. The vocal tract filter is determined by the size and shape of the vocal tract and is therefore directly affected by the position and movement of the articulators such as the tongue, jaw, and lips.

Vowels are typically characterized in terms of the location of the first two formants, as illustrated in Figure 1 for the vowels of American English. For a given speaker, each vowel typically has a unique formant pattern. However, variation in vocal tract size among speakers often leads to a degree of formant overlap for different vowels.

Consonants can also be described in terms of their spectral properties. These sounds are produced with a complete or narrow constriction in the vocal tract, essentially creating a vocal tract with two sections: one behind and the other in front of the constriction. The length of the section in front of the constriction is one of the primary determinants of the spectra of these sounds. The longer this section (i.e. the farther back the constriction), the lower the frequency at which a concentration of energy occurs. For example, consonants like *k* and *g,* which are produced at the back of the mouth, are typically characterized by a concentration of energy between approximately 1,500 and 2,500 Hz, whereas more anterior consonants like *t* and *d* typically have a concentration of energy above 3,000 Hz. Similarly, the sibilants [ʃ, ʒ] produced in the middle of the mouth have major energy around 2,500 to 3,500 Hz, whereas the more anterior ones [s, z] have major energy well above 4,000 to 5,000 Hz. However, in the

case of consonants with a constriction toward the very front of the vocal tract, the extremely short section in front of the constriction does not result in clearly defined spectra. As a result, bilabial [b, p] and labiodental [f, v] consonants are described as having diffuse spectra, without any clear concentration of energy.

From a linguistic point of view, a detailed description of speech sounds in terms of their frequency, in addition to amplitude and duration, can elucidate the factors that shape sound categories and determine phonological processes both within and across languages. In addition, acoustic phonetic analysis may serve to quantify atypical speech patterns produced by nonnative speakers or speakers with specific speech disorders.

References

Asher, R., and Eugenie Henderson (eds.) 1981. Towards a history of phonetics. Edinburgh: Edinburgh University Press.

Fant, Gunnar. 1960. Acoustic theory of speech production. The Hague: Mouton.

Flanagan, James L. 1965. Speech analysis synthesis and perception. Berlin: Springer-Verlag.

Hardcastle, William, and John Laver (eds.) 1997. The handbook of phonetic sciences. Oxford and Cambridge, MA: Blackwell.

Helmholtz, Hermann von. 1877. Die Lehre von den Tonempfindungen als physiologische Grundlage fur die Theorie der Musik. Braunschweig: F. Vieweg; as On the sensations of tone as a physiological basis for the theory of music, translated by Alexander J. Ellis, London: Longmans, Green, 1885.

Hillenbrand, James, Laura A. Getty, Michael J. Clark, and Kimberlee Wheeler. 1995. Acoustic characteristics of American English vowels. Journal of the Acoustical Society of America 97(5). 3099–111.

Jakobson, Roman, Gunnar Fant, and Morris Halle (eds.) 1952. Preliminaries to speech analysis. Cambridge, MA: MIT Press.

Johnson, Keith. 1997. Acoustic and auditory phonetics. Oxford and Cambridge, MA: Blackwell.

Kent, Raymond D., Bishnu S. Atal, and Joanne L. Miller (eds.) 1991. Papers in Speech Communication: speech production. Woodbury, New York: Acoustical Society of America.

Ladefoged, Peter. 1962. Elements of acoustic phonetics. Chicago and London: The University of Chicago Press, 2nd edition, 1996.

Lehiste, Ilse. 1970. Suprasegmentals. Cambridge, MA: MIT Press.

Lieberman, Philip, and Sheila E. Blumstein. 1988. Speech physiology, speech perception, and acoustic phonetics. Cambridge: Cambridge University Press.

Stevens, Kenneth N. 1998. Acoustic phonetics. Cambridge, MA: MIT Press.

ALLARD JONGMAN

Acquisition

Language acquisition is the study of the development of a person's language. It generally refers to the way people learn their native, first, second, or other languages. More specifically, it may refer to the time a language feature has been acquired. This may vary from the first emergence or onset of a language item to the time of its accurate use. As a field of study, it is the subject of linguistics, psychology, and applied linguistics. Its object is to study (1) how languages are learned, (2) what are the developmental stages in this process, and (3) what is the nature of language. To find answers to these questions, researchers apply longitudinal and cross-sectional methods. In the first of these, they study specific developments in the language of individuals or groups over a period of time. In the second, they research a particular feature in the language of a group at a given point in time.

First Language Acquisition

First language acquisition is the child's learning of his or her first or native language. Traditionally, and especially in monolingual societies, 'first' and 'native' language were used synonymously. With the expansion of cross-cultural communication, the two terms become more distinct. For example, children may acquire some knowledge of another language from a nurse or a relative before they acquire their native language, e.g. the language of the country they live in. Thus, a Chinese child born in the United States may first learn Chinese from her parents, and learn English later from English-speaking children and adults. To avoid the confusion arising from the use of 'first' and 'native', another term, 'primary', is sometimes used to indicate a child's first language chronologically.

First Language Acquisition and the Language Acquisition Device
Noam Chomsky's work aroused interest in the way children learn their native language. He believes that children are born with the ability to learn a language, i.e. they are born with a 'language acquisi-tion device'. The latter is species-specific or only for humans, language-specific or only for their first/native language, and innate or only inborn. He also claims that this ability is unconscious and children learn their native language by exposure to it and by using it, and not by being taught or corrected. He argues that as children acquire their native language, they are able to produce sentences that they have not heard before.

While early work on children's language acquisition focused on the development of children's ability to produce novel sentences, more recently, researchers have emphasized children's acquisition of word meanings and their linguistic and cognitive development, their acquisition of the phonology of their native language, and their language development in relation to their interaction with parents and peers. Some researchers also see a parallel between the stages of children's language development regardless of the specific language they are learning.

First Language Acquisition and Cognitive Development
A child's language development is closely related to his or her cognitive development. Here, the ability to identify and form categories and concepts is of crucial importance. 'Categorization' involves the treatment of distinct linguistic phenomena, such as 'worked', 'studied', 'saw', and 'went', as if they were part of the same phenomenon, or the same grammatical category, i.e. past tense. Young children do not have fully developed abilities in categorization. Many childhood first language errors, for example, '*I eated it', point to the gaps in their ability to form categories. Furthermore, even seemingly correct utterances do not imply that the child has achieved an adult stage in the mastery of the corresponding language category.

3

Closely related to the ability to categorize is the ability to differentiate a category, for example, tense, from the mental structure, which it represents, in this case, time. These mental structures are known as 'concepts'. To learn a language, a child must acquire the concepts that underlie linguistic structures. It is not possible to master grammatical categories, such as tense, in any language without mastering concepts such as time, space, modality, causality, and number. Young children's errors in tense indicate that they do not grasp the concept of time.

First Language Acquisition and Social Development

Children's social adjustment is as important as their cognitive growth to their language development. As they acquire various language categories and the concepts they represent, children also learn about the cultural, moral, religious, and other conventions of the society they live in. They learn how to express their thoughts, feelings, and wishes in a socially acceptable manner. For example, children learn that it is not always advisable to speak their minds. As they come to realize that words can serve to make friends as well as enemies, they learn that it is not always possible to tell the truth. In this way, while acquiring their first language, children also develop a social identity. How-ever, their progress is slow and not devoid of some rather amusing or even embarrassing errors. Children learning their first language, therefore, have a long way to go, even after they have acquired the basic concepts and their corresponding language categories. By comparison, adult foreign language learners, who are knowledgeable about the sociocultural aspects of their native language, are a step ahead of child learners, even though they may also be prone to similar social blunders because of sociocultural differences.

First Language Acquisition and the Critical Period

The 'critical period hypothesis' claims that there is a period in child development during which language can be acquired with native-like proficiency. Some, like the biologist Lenneberg, believe that this period lasts until puberty, after which the brain loses some of its ability to adapt due to its laterization, i.e. the establishment of specific language functions in particular parts of the brain. After that, the decreased plasticity of the brain makes the acquisition of another language a psychologically different and more difficult process. While there is compelling evidence that supports those claims, there are also important facts that undermine their veracity. First, the strictly biological evidence is by no means conclusive. Second, other factors, such as lack of motivation, may explain nonnative pronunciation.

First Language Acquisition and Bilingualism

Cognitive and social development, as well as the language acquisition device and the critical period affect the language development of the bilingual as well as the monolingual child.

Bilingual first language acquisition is defined as the parallel acquisition of two languages, which is, supposedly, an evenly paced process. However, such a perfect balance can rarely be achieved. Commonly, the child would use one language in one environment, and another in a different setting. Thus, inevitably, one language gains dominance over the other. This dominance may extend to some or all areas of communication. As a result, the child's other language may become secondary in both development and use. Furthermore, there may be some interference from the dominant language that causes errors in the child's secondary language. However, there is no evidence that this results in massive confusion in one or both languages. Furthermore, there is no evidence that bilingual children differ from other children in their cognitive, social, or language development.

Second Language Acquisition

Second language acquisition (SLA) is defined as the process of becoming competent or proficient in a second or foreign language, from the first use of a language item to its advanced applications at a later stage. As a field of research, SLA is a fairly new interdisciplinary subject with most of its empirical research done since the 1960s. It is largely based on theories and research methods developed in the fields of education, psychology, linguistics, anthropology, foreign languages, English as a second language, and linguistics. In the United States, researchers study the way nonnative speakers acquire English phonology, syntax, and pragmatics. The purpose of SLA studies is to describe and explain the way second languages are learned in terms of both linguistic and communicative competence. To do this, researchers study learners' performance and their intuitions about correct and incorrect use of language. The object of second language acquisition is to find more effective ways of teaching and learning foreign languages, and assumes that such research can affect the way foreign languages are learned.

The Meaning of 'Second Language' in Second Language Acquisition

There are different interpretations and uses of both 'second' and 'acquisition' in SLA. 'Second' may be used to distinguish it from 'foreign' or 'third' language acquisition. Traditionally, the terms 'second' and 'foreign' have been used alternatively to refer to any language other

than the first. More recently, with the emergence of English as a global language and the establishment of Teachers of English to Speakers of other Languages (TESOL) as a worldwide professional organization, a distinction is made between the two. 'Second' language acquisition refers to the study of English by foreigners in countries where English is the native or the official language, whereas 'foreign' language acquisition refers to the study of English everywhere else. Furthermore, this distinction extends to differences in what is learned and how it is learned. Learners of English as a foreign language (EFL) prefer standard varieties of English, whereas learners of English as a second language (ESL) try to blend with their sociolinguistic environment. All of these differences are reflected in the goals and methods of EFL and ESL. A further distinction is made between 'second' and 'third' language acquisition, which marks the learner's relative proficiency rather than the order in which he or she acquired these languages. Sometimes, the term 'alternative' is used to refer to any nonnative language.

The Meaning of 'Acquisition' in Second Language Acquisition

Acquisition is often used to refer to different aspects of the process and results of learning a second language. While trying to find out about the process, i.e. how second languages are learned, researchers often compare different learning experiences that lead to SLA, such as learning a language through organized instruction or in an immersion situation. From a sociolinguistic perspective, acquisition through organized instruction occurs in classrooms with the help of teachers and instructional materials. Acquisition through immersion occurs in social situations using contextual clues. Yet another distinction is made from the psycholinguistic perspective. Klein identifies 'spontaneous' and 'guided' acquisition. The first focuses on everyday communication, whereas the second targets the mastery of the language system. Similarly, from a psychological point of view, Krashen distinguishes between 'acquisition' and 'learning'. In his analysis of the process of mastering a second language, he reserves 'acquisition' for the subconscious process of learning a language by being exposed to it. 'Learning', according to him, is the conscious process of mastering a language by studying it. Ellis finds this distinction problematic and considers its demonstration difficult. Furthermore, he states that researchers disagree about what kinds of performance constitute evidence of 'acquisition'. For some, such evidence can be found in the ways learners speak and write. For others, it is the learner's intuitions about the second language that matter. Yet another group of scholars seek evidence of acquisition by assessing the learner's introspections.

Other researchers analyze what it means to know a second language. From a linguistic perspective, Chomsky focuses on the results of SLA, which he defines as 'competence' and 'performance'. According to him, 'competence' in a second language is the mastery of the internalized grammar that the ideal speaker or hearer, not a real one, has of the whole language. Such mastery enables him/her to produce grammatically correct sentences as well as recognize existing and nonexisting sentences. For example, knowing the rule, which makes 'I speak English' possible, a person can produce 'I speak French' even though he or she may not have seen this before. Furthermore, he or she would know that the form '*I speaking English' is nonexistent. 'Performance' in Chomsky's *Generative Transformational Grammar*, on the other hand, refers to a person's actual use of a language in the understanding and production of sentences. Unlike communicative competence, which is internal and invisible, performance is external and observable. Furthermore, performance does not mirror competence, since people may know how to produce a sentence but may err when they try to do so. Thus, performance can also be defined as the grammar that a person uses to understand and produce language, which is both correct and appropriate. Performance could be used to investigate competence through the analyses of samples of spoken or written discourse. Within performance, Widdowson distinguishes 'usage', which refers to the learner's ability to apply grammar rules accurately in the production of grammatically correct sentences. 'Use', on the other hand, signifies their ability to apply linguistic and sociolinguistic knowledge appropriately and communicate effectively in diverse contexts.

Second Language Acquisition Research

A large part of SLA research is learner based. It describes and analyzes the nature of learner language and learner differences, learning processes, and pedagogical input and output. It does so to provide answers to important second language research questions, which may offer effective solutions to crucial second language classroom problems.

Learner Language in Second Language Acquisition Research

Researchers study learner language by examining samples of oral and written texts. Their goal is to identify errors, establish developmental patterns and sequences, trace variability, and explore use. Errors were first believed to be the result of native language 'transfer' or 'interference'. This view was promoted by numerous contrastive analyses conducted from the 1940s to the 1960s. Such studies compared two languages to find out what similarities and differences

existed between the two. Lado thought that similar elements would be easy to learn, while dissimilar ones would be difficult to master. The belief that linguistic difference could be a predictor of difficulty gave rise to the 'contrastive analysis hypothesis'. That and the behaviorist approach to learning, which claimed that learning is a process of habit formation, led to the belief that SLA should be a process of overcoming habits from the native language and consolidating correct habits in the target language.

In 1967, Pitt Corder proposed a new definition of errors. He thought they were systematic deviations from the norm, which reflect the learner's current stage of second language development. Errors, he claimed, are different from mistakes, which can easily be self-corrected. 'Error analysis' treated errors as a sign of the learner's hypothesis testing, which would ultimately lead to the formation of the correct form and its underlying rule. Thus, errors were seen as part of the learner's language at every stage of its development. To emphasize its unique features, Corder referred to learner language as 'idiosyncratic dialect'. Nemser called it an 'approximative system' and Selinker coined the term 'interlanguage'. Thus, the notion of learner language evolved from a faulty, deviant product of the target language to a continuous, approximative progress towards its mastery. Both contrastive and error analysis were criticized for their exclusive reliance on the analysis of a linguistic product, i.e. errors, to yield insights into a psycholinguistic process, i.e. second language acquisition.

Another feature of learner language is its passage through a sequence of developmental stages, which are universal. Thus, many of the initial utterances that learners produce may be simple formulae, for example, 'What's this'. These are followed by structures of greater morphological and syntactic complexity, for example, 'I wonder what this might be'. The existence of developmental stages in SLA, which are similar to those in first language acquisition, along with some variations in the specific order in which particular features occur, have renewed interest in grammar instruction. In its systematic development, learner language also exhibits certain variability. For example, learners may use the third person singular 's' correctly sometimes and omit it at other times. In addition to lexical and syntactic variability, they often have problems on the pragmatic level, i.e. they may use language or act in a socially inappropriate manner.

Factors in Second Language Acquisition

SLA is also influenced by the environment in which it occurs. Social factors, language input, and interaction affect the way learners acquire a second language. Ellis contends that social factors shape learners' attitudes, which, in turn, may affect motivation and learning outcomes. Social factors include natural and educational settings. For example, natural settings, where English is the native or the official language, offer opportunities different from educational settings, such as the foreign language classroom where the native language is the medium of instruction.

While social factors influence second language acquisition indirectly, input, output, and interaction seem to have a direct impact. 'Input' is the learner's access and exposure to the second language, both written and oral. Exposure to the foreign language may engage learners in 'interactions' with native or nonnative speakers, or it may involve them in listening to tapes, films, radio, and TV programs. Researchers vary in their assessment of the importance of input and interaction. Behaviorist theories emphasize the importance of input. Chomsky, on the other hand, claims that there is no necessary correlation between language input and learner output. Krashen believes that learners acquire language in a natural order as a result of being exposed to 'comprehensible input' addressed to them. In contrast to Krashen, Swain proposes the 'comprehensible output hypothesis', which claims that comprehending input alone will not prepare students to produce language. According to him, it is correct production resulting from challenging practice in speaking and writing that facilitates acquisition. Both the comprehensible input and comprehensible output hypotheses have been criticized on the grounds that the processes of comprehensible input and output and the process of SLA are not the same.

General factors, such as social setting, input, output, and interaction, result in a variety of individual differences in SLA. Furthermore, individual factors, such as age, language aptitude, motivation, cognitive style, and learning strategies can have similar effects. These factors affect second language learning in ways that are mostly independent of the learner. For example, a learner can do nothing about his or her age, or language aptitude. Few learners may have the opportunity to switch from one educational setting to another. Given the appropriate guidance, however, some learners may be able to improve their motivation and learning strategies over time. For example, 'extrinsic motivation', which derives from external rewards, may evolve into 'intrinsic motivation', which derives from personal interests. Learner strategies, which contribute to the learner's conscious efforts to learn, may also change. For example, learners may expand their 'cognitive strategies' by learning new concepts. They may also perfect their 'metacognitive strategies' by

developing their study skills, or enhance their 'social strategies' by practicing their knowledge in authentic social settings.

References

Bennet-Castor, T. 1988. Analyzing children's language. Oxford: Blackwell.

Brown, Douglas H., and Susan Gonzo. 1995. Readings on second language acquisition. Englewood Cliffs, NJ: Prentice-Hall Regents.

Chomsky, Noam. 1965. Aspects of the theory of syntax. Cambridge, MA: MIT Press.

Cook, Vivian. 1993. Linguistics and second language acquisition. New York: St. Martin's Press.

Ellis, Rod. 1994. The study of second language acquisition. Oxford: Oxford University Press.

Fletcher, P. 1985. A child's learning of English. Oxford: Blackwell.

Fletcher, P., and M. Garman (eds.) 1979. Language acquisition. Cambridge: Cambridge University Press.

Hatch, E. (ed.) 1978. Second language acquisition. Rowley: Newbury House.

Klein, Wolfgang. 1986. Second language acquisition. Avon: Cambridge University Press.

Krashen, Steven. 1981. Second language acquisition and second language learning. Oxford: Pergamon.

Larsen-Freeman, Diane, and Michael Long. 1991. An introduction to second language acquisition research. London and New York: Longman.

Lenneberg, E. 1967. Biological foundations of language. New York: Wiley.

Long, Michael. 1995. The least a second language theory needs to explain. Readings on second language acquisition, ed. by Douglas H. Brown and Susan Gonzo. Englewood Cliffs, NJ: Prentice-Hall Regents.

Richards, Jack, et al. 1992. Longman dictionary of applied linguistics. Harlow, Essex: Longman.

Widdowson, Henry. 1978. Teaching language as communication. Oxford: Oxford University Press.

Lilia Savova

See also **Acquisition Theories**

Acquisition Theories

The goal of acquisition theories is to explain how it is that any normal child, born into any linguistic community, learns the language (or languages) of that community. For many theorists, the challenge is also to explain what appears to be the relatively short time period in which acquisition is achieved, the fact that it appears to be done without either overt teaching or sufficient information from the input (what the child hears [or sees, in the case of sign language]), and to follow a path that seems remarkably similar in all children, despite variation in early childhood experiences and in the types of languages they are exposed to. There is also a consensus that language acquisition is largely independent of cognitive development, despite the fact that some deficits in cognitive development can have an effect on certain aspects of language development. Whether the language is a spoken language or a sign language, whether the language is highly inflected like Finnish or uninflected like Mandarin, whether the child is raised in poverty or luxury, by highly educated or illiterate adults, or even other children, it seems that normally developing children pass through roughly the same stages in the same sequence, and achieve the steady state of acquired language by about the same age.

We know from unfortunate natural experiments in which children are raised in isolation (or near isolation) from language-using older members of our species (i.e. they are severely neglected, or are raised by other animals) that language does not emerge without at least some linguistic exposure—input—during the first few years of life. This strongly suggests that there is a critical or sensitive period during which the mechanism or mechanisms responsible for language development is/are primed to receive input. However, the resilience of language development to quite wide variations in input within any given language community, as well as the similarities among children learning quite different languages, suggest that these mechanisms, whatever they are, must either be quite tolerant of such variation, or be primed in such a way that the crucial input for language acquisition is always made available.

There are important roles for both 'experience-dependent' (nurture) and 'experience-expectant' (nature) learning in language development, and theories are distinguishable in terms of the relative contributions they see for these two types. On the one hand, there are those researchers who see a large didactic role for input (experience-dependent learning), and on the other, there are those who see a much smaller role for input and a much larger role for genetic predispositions that are triggered by linguistic experience (experience-expectant learning). Theories are also distinguishable

in terms of whether they are trying to explain how language emerges on a day-by-day basis in any given child or whether they are trying to explain how what is perceived as a gulf between experience as a child and adult knowledge of language could be bridged in principle. The latter are engaged in trying to solve the 'logical problem of language acquisition'.

Another dimension of difference between theories concerns the nature of the experience-expectant (innate) aspects of language acquisition. There are those researchers, most notably the generative linguists in a broadly Chomskyan paradigm, who argue for a dedicated language acquisition device (LAD), that has evolved to serve the precise purpose of language acquisition. This device is primed specifically to receive linguistic input, and requires a minimal amount of it to set to work building mental representations for language in the mind of the child. The fact that the required triggering input is so minimal provides an explanation for the consistency of language acquisition paths across otherwise fairly widely varying life experiences.

The most elaborated version of the LAD account—the Principles and Parameters approach—suggests that children are born with a Universal Grammar (UG), which means they are (unconsciously) anticipating those features that are common to all languages (the principles), as well as limited options for those things that differ among languages (the parameters). Upon exposure to actual input from a given language, children are able to 'decide' which sort of language they have encountered. So, for example, some languages have basic subject–verb–object organization in which complements are attached to the right of the heads of phrases (thus objects follow verbs, relative clauses follow noun heads, and noun phrases follow prepositions), while other languages are subject–object–verb where the reverse order of complements is found. A child exposed to a language of the first type need only process a simple structure (say one with a verb followed by an object) and it will trigger the expectation that all the other head-complement structures will be in the same order. When all the open parameters have been set, the child possesses the 'core' grammar of the specific language he or she is exposed to. At the same time, however, the child has also been acquiring those aspects of the language that are not anticipated by UG, using experience-dependent learning. These aspects are often together referred to as 'peripheral' grammar. Some researchers in this paradigm have assumed that all principles and parameters are operational or sensitive to the input from the beginning of life. Others have suggested that some may at least emerge with maturation.

In the course of acquisition, generalization (and overgeneralization) allows new knowledge to permeate across-the-board, and 'bootstrapping' allows learning in one area of language to trigger new learning in another. Semantic bootstrapping involves understanding an utterance in context and using it in conjunction with innate expectations of language principles to crack the code of the syntax. (For example, a child who does not yet know the required complements of the verb 'put', can work these out from understanding utterances such as 'Put the cup on the table' in context.) Syntactic bootstrapping involves working from an already understood structure to fill in meanings, semantic information, by deduction. (For example, if you hear 'John glopped his friend on the head and he fell down', you may not know exactly what 'glopped' means, but you can work out a lot of what it must mean.)

The Principles and Parameters (P&P) model has been a highly influential linguistic approach in language acquisition research, even while researchers in psychology and anthropology have been pursuing significantly different lines of investigation. In linguistics, the P&P model lies at the intersection of generative (specifically Chomksyan) linguistic theorizing about the nature of adult mental representations for language and accounts of how children acquire language. It has evolved as an account of how language could actually develop across time, even while its roots are in the 'logical problem of language acquisition' because it assumes that what cannot be learned from the input must be genetically prespecified. It advances various arguments in support of the thesis that the input is in fact incapable of providing sufficient information for language to be learned entirely through experience-dependent mechanisms, and thus that there is an essential problem of the 'poverty of the stimulus'. The account is also strongly modular in the sense that it sees language as being acquired by a specially dedicated mechanism (the LAD), independent of other types of mental representations or mental functioning.

Despite its power within linguistics, accounts predicated on some version of the UG story actually attract only a minority of adherents within the broad field of child language research. Most researchers are convinced that language in its entirety can be worked out by the child on the basis of the input, coupled with innate (nonspecific) predispositions for analyzing their environment. As a result of their analysis, children possess the capacity to produce and comprehend language. They may also have stored mental representations for how language works, but, unlike the P&P account, most of the accounts that rely on processing of the input as the way in which acquisition occurs also regard the endpoint as processes for production

and comprehension, without independent mental representations. Ann Peters and Dan Slobin, for example, have argued that children possess strategies, operating principles, for carving up the speech stream into repeating bits, storing those bits, recognizing the commonalities across them, and thereby building up a performance grammar for comprehension and production called Basic Child Grammar, which subsequently becomes elaborated into an adult grammar.

Another of the processing theories of language acquisition is Elizabeth Bates' and Brian Mac Whinney's Competition Model. Unlike the operating principles approach, this does not argue for a steady state in the form of a grammar at the end of acquisition, but rather for a permanently dynamic response to input throughout life. The impression that language has been definitively learned comes simply from the fact that new input changes the child's system very little if a child remains in the same speech community (although it will usually change with exposure to a new dialect or to a new language). Acquisition takes place as children respond to the distribution of various cues to meaning (word order, inflectional information, intonation and stress, etc.) and respond probabalistically to conflict among them. Since each child processes the input independently, individual differences between children are expected, and advocates of this approach argue that the differences found among children support the model. As they learn more of their language, they pay attention to more and more of the cues and let the stronger ones win out over the weaker ones for their language. So, at an early stage, a child learning English may assume that nouns at the beginning of sentences are agents, but when they begin to pay attention to passive morphology, they will have to adjust their assumptions accordingly. As should be clear from this example, the ability to derive meanings for language from context, in advance of actually understanding how language is structured is crucial (as it is in semantic bootstrapping). Active application of this and other distributional models, aimed at demonstrating that the input is sufficient to account for language acquisition, is seen in the computer-modeled connectionist networks. Attracted by the architectural similarity between computer networks and neurons, these researchers argue that experience with language teaches the network so effectively and so quickly that it gives the impression of prior knowledge.

As already indicated, children frequently seem to rely on pragmatic expectations of what language ought to mean in a given situation as a basis for learning how it is structured. Some theories of language acquisition place an even greater emphasis on the role of the sociopragmatic environment. Jerome Bruner,

for example, argues that the behavior of a child's caretaker provides all the cues that the inherently social child needs to acquire language. The way in which language is used in the here and now, in conjunction with actions that match what is being said, 'scaffolds' the child's understanding of the language. Others see difficulties with this approach, not only because not all successful language learners receive the kind of careful scaffolding it seems to require but also because the parental complexity of language seems to follow rather than lead increasing complexity in the child's language. In these and other ways, the account seems too simple to account for the complexity of the task. A similar approach was advocated by Jean Piaget, who saw language development as the logical extension not of social behavior, but of cognitive development. In Piaget's initial proposal, embedding of sentences was seen as analogous to nesting boxes, and the former dependent on the latter. Although most of his specific predictions have not been supported by subsequent research, it is clear that at least certain aspects of language are intimately related to cognitive development, although the direction of influence is not clear. One area of current research concerns the emergence of the capacity to make informed guesses about what other people know ('theory of mind'), and its relationship to language development. There is considerable discussion about whether the capacity to embed clauses under main verbs such as 'think' and 'know' drives or is driven by the capacity to understand that others may not share the same assumptions as oneself.

Finally, it is worth noting that significant debate surrounds the issue of whether, when an individual learns a second language, they go about it in the same way as a first language learner, or whether it is a fundamentally different process. Evidence suggests that there is some kind of critical period for second language development, as children seem to be able to do it so much better than adolescents and adults. However, it is not yet known when bilingual first language development should be seen as having given way to second language development. Nor is it clear that adults, when given the kind of input and motivation of children, are always incapable of the same level of success. It is also unclear whether second language learners are able to reaccess the learning capacities they had as children learning their first language, or whether these are permanently overridden and made unavailable by the presence of the first language.

References

Crain, Stephen, and Diane Lillo-Martin. 1999. An introduction to linguistic theory and language acquisition. Oxford: Blackwell.

de Boysson-Bardies, Benedicte. 1999. How language comes to children. Cambridge, MA: MIT Press.

Fletcher, Paul, and Brian MacWhinney (eds.) 1997. The handbook of child language. Oxford: Blackwell.

Foster-Cohen, Susan. 1999. An introduction to child language development. Harlow, Essex: Longman/Pearson Education.

Hoff-Ginsberg, Erika. 2000. Language Development, 2nd edition. Pacific Grove, CA: Brooks/Cole Publishing Company.

Jackendoff, Ray. 1995. Patterns in the mind: language and human nature. New York: Basic Books.

Karmiloff, Kyra and Annette Karmiloff-Smith. 2001. Pathways to language: from fetus to adolescent. Cambridge, MA: Harvard University Press.

Owens, Robert. 1996. Language development: an introduction, 4th edition. Boston: Allyn and Bacon.

Pinker, Steven. 1994. The language instinct: how the mind creates language. New York: William Morrow and Company Inc.

Ritchie, William and Tej Bhatia (eds.) 1999. Handbook of child language acquisition. New York: Academic Press.

Smith, Neil. 1999. Chomsky: ideas and ideals. Cambridge: Cambridge University Press.

SUSAN H. FOSTER-COHEN

See also **Acquisition; Developmental Stages; Slobin, Dan Isaac**

Aerodynamics of the Vocal Tract

Speech production may be viewed as a means of converting slow pressure variations in the vocal tract into the rapid pressure variations that constitute sound. When there is a difference in air pressure across a valve and it opens, sound is created by the rapid movement of air. A stop burst is created by the sudden release of air under pressure, the continuous noise of a fricative is created by a similar but slower release, and repetitive release of air through the glottis by the vibrating vocal cords creates voicing. Speech aerodynamics studies in detail as to how these sounds are created, how the air pressures and airflows are generated, and their phonological consequences.

From an aerodynamic point of view, the vocal tract is a system of chambers connected to each other by valves, with some of them having access to the air outside. Most of the chambers are equipped with piston-like structures that can change the volume of their chamber.

There are a few general physical principles that govern the generation of pressures and flows in the vocal tract:

1. For a given mass of air, pressure varies inversely with the volume of the chamber containing it (Boyle–Mariotte's Law).
2. For a given volume of a chamber, pressure varies directly with the mass of air in it.
3. The quantity of airflow (the 'volume velocity') through an aperture varies as a function of the size of the aperture and the pressure difference across it.
4. The velocity of airflow (the so-called 'particle velocity') varies directly with the volume velocity and inversely with the size of the channel it is flowing through. (This principle is important because turbulence, the source of acoustic noise, is a function of air velocity.)
5. When air is flowing through a channel, the pressure it exerts on the walls of a channel at 90° with respect to the direction of airflow varies inversely with the velocity of the airflow (this is the Bernoulli effect).

Four 'air stream mechanisms' for generating the pressures needed in speech are used in languages of the world. They are named for the principal piston-like structure creating the volume change, plus 'egressive' or 'ingressive' to denote whether air moves out of the cavity or into it when the pressure is released to the atmosphere:

Pulmonic egressives: The air in the lungs is compressed by the expiratory muscles acting to decrease the volume of the rib cage; the pressure increases (by principle 1). Sound is created when this air is released at the glottis and/or at any supraglottal constriction, e.g. [p], [l], [m], [o]. All languages use pulmonic egressive sounds and the vast majority like English, Japanese, Finnish, and Hawai'ian use them exclusively.

Glottalic egressive ('ejectives'): With air trapped in the oral cavity between a closed glottis and a supraglottal constriction (and, of course, velic closure), the air is compressed by the elevation of the larynx (the glottis acting like a piston) and some pharyngeal constriction. Such sounds may include stops, affricates, and fricatives, e.g. [p'], [ts'], [f']. Ejectives are found in approximately 18% of the world's languages, e.g. Amharic, Quechua, and Navaho.

Glottalic ingressive ('implosives'): The oral cavity is enlarged (by lowering the larynx and possibly also the jaw) so that the pressure in the oral cavity decreases. Most implosives are voiced, the lowering larynx thus acting like a leaky piston, but voiceless implosives, although quite rare, do exist. Voiced implosives occur in about 10% of the world's languages, e.g. Kalabari, Sindhi, and Xhosa.

Velaric ingressive ('clicks'): Air is trapped between the tongue and the palate (or the lips) and the small cavity is enlarged by the tongue moving down and/or backwards. The air pressure is greatly lowered and when released by the tongue or lips, producing a 'pop' or fricative sound. The sound that comic strips try to convey with 'tsk'! or 'tut-tut' (a sound expressing disapproval or regret) is the alveolar click [!]. The sound of a 'kiss' is the bilabial click [ʘ]. Clicks are found widely in different cultures, but usually as interjections or calls to animals. As speech sounds, they are quite rare and are found only in certain Southern African languages such as Zulu, Xhosa, Ndebele, and !Xóõ. Given that clicks' sound-making structures are so localized in one part of the vocal tract, clicks can be combined simultaneously with sounds made by the other air stream mechanisms, e.g. they can be accompanied by voicing or voicelessness, nasalization, etc.

Voicing and trills: Trills and voicing constitute aerodynamically driven, self-sustaining oscillations. The cycle of vocal cord vibration could start with the vocal cords lightly adducted in the presence of airflow. This causes the subglottal pressure to build up behind the glottal closure, the increased pressure forces the cords apart, and the resulting high rate of airflow through the cords creates a negative pressure between them (by principle 5, above), which 'sucks' the cords together again (along with an elastic recoil force of the vocal cord tissue); the cords close again, and the cycle repeats itself. Trills are created in a similar way.

Obstruent and voicing: Obstruents inherently favor voicelessness for the following reason: voicing requires that there be air flowing between the vocal cords. Obstruents partially or completely block the airflow exiting the oral cavity. But if glottal airflow continues, then the pressure in the oral cavity will rise (by principle 2, above). Then, the difference in the pressure across the glottis will be reduced and may fall below the value needed to maintain the airflow required for vocal cord vibration (by principles 3 and 5, above). Passive yielding of the vocal tract walls to the impinging pressure permits voicing to persist for around 65 msec (more for labial stops, which have more surface area behind the constriction and less for velars). For voicing to be maintained beyond this, some active expansion of the oral cavity is required, e.g. by lowering the larynx and/or the jaw. This leads to many phonological consequences, among them the greater incidence of 'missing [●]' in the stop inventories of languages utilizing a voicing contrast on stops, e.g. as in Dutch, Czech, and Thai, the greater tendency for geminate stops to be voiceless or to undergo devoicing morphophonemically.

Many other common sound patterns in language, e.g. the development of affricates from stops, as well as the way specific speech sounds are modified in certain cases of speech disorders, e.g. cleft palate, laryngectomy, can be explained by reference to speech aerodynamics.

References

Ladefoged, P., and I. Maddieson. 1996. The sounds of the world's languages. Oxford: Blackwell.

Ohala, J.J. 1983. The origin of sound patterns in vocal tract constraints. The production of speech, ed. by P.F. MacNeilage, 189–216. New York: Springer-Verlag.

Ohala, J. J. 1997. Aerodynamics of phonology. Proceedings of the 4th Seoul International Conference on Linguistics [SICOL], 92–7. 11–15 August 1997.

JOHN J. OHALA

See also **Speech Production**

Affixation

Affixation is a morphological process that adds phonological material to a word in order to change its meaning, syntactic properties, or both. Some examples of affixation in English are given in (1).

(1) a. fond   fondness

    b. start   restart

    c. car   cars

In this article, an overview of different types of affixation processes is given, followed by a discussion of how to distinguish affixation from other processes. Then, a number of theoretical problems related to affixation are presented.

Types of Affixation

The phonological material added in affixation is called the *affix*. The affix is attached to the *base*. Different types of affixation can be distinguished on the basis of the position of the affix with respect to the base or on the basis of how affixation affects the meaning of the base. In this section, the emphasis is on the former. The latter is discussed in the section on delimitation issues.

Suffixation is the most common type of affixation. In suffixation, the affix is added to the end of the base. Examples are (1a) and (1c) above. In (1a), the suffix-*ness* is added to the adjective *fond* to produce the noun *fondness*. In (1c), the suffix -*s* is added to the noun *car* to produce the plural of the noun. In most languages, suffixation is the most widespread form of affixation. In languages such as Turkish and Finnish, it is the only type of affixation.

In many respects, *prefixation* is the mirror image of suffixation. In prefixation, the affix is attached to the beginning of the base. An example is (1b) above, where the prefix *re-* is added to the base *start*. Prefixation is less widespread than suffixation, but some languages, e.g. Khmer, only have prefixation. In languages that have both suffixation and prefixation, the former usually has a larger range of functions. Thus, in English, all inflection is expressed by suffixation. In word formation, we find both, but prefixation only rarely changes the syntactic category of the base, as when the verb *enrich* is formed from the adjective *rich*.

In most languages, affixation only involves suffixes and prefixes. *Infixation*, in which the affix is attached inside the base, is much rarer. An example is the infix -*um*- in Tagalog, illustrated in (2).

| (2) | a. bilih ('buy') | bumilih ('bought') |
| | b. gradwet ('graduate') | grumadwet ('graduated') |

The infix in (2) is used to express the past tense of the verb. It attaches itself after the first consonant or cluster of consonants of the base. This is its *anchor point*. In general, the anchor point of an infix is at most one syllable removed from the left or right boundary of the base.

Whether other types of affixation exist depends on the theoretical framework adopted. Thus, *circumfixation* is the simultaneous addition of a prefix and a suffix. The mere observation of a form such as *enrich-*

ment is not sufficient evidence for the existence of a circumfix in English, because the form can be analyzed transparently as a result of suffixation of -*ment* to the verbal base *enrich*, which is in turn the result of prefixation of *en-* to the adjective *rich*. A better candidate for a circumfix is Dutch *ge…te,* exemplified in (3).

| (3) | steen ('stone, rock') | gesteente ('type of stone/rock') |

The change in meaning in (3) can be expressed as a generalization to a type. It can be argued that *ge…te* is a circumfix, because neither *gesteen nor *steente are possible Dutch words. A theory that excludes circumfixes would have to postulate one of these nonexisting forms as an intermediate form in the derivation.

Delimitation Issues

As affixation is the most salient morphological process, it is no surprise that many of the foundational discussions in morphology are directly related to affixation.

A first set of issues concerns the function of affixation. Traditionally, inflection and derivation are distinguished. Inflection adapts a word to its syntactic environment. A prototypical example is the agreement of a verb with its subject in person and number, as illustrated for *sleep* in (4).

| (4) | a. The child sleeps. |
| | b. The children sleep. |

Derivation is a type of word formation that involves affixation. It forms new lexemes, marked by a different syntactic category and/or a different lexical meaning, as illustrated in (1a-b). Another contrast is the one between inflection and stem formation. An example of the latter can be observed in Dutch *kinderstoel* ('highchair'), a compound of *kind* ('child') and *stoel* ('chair'). The form *kinder* is different from the singular (*kind*) and the plural (*kinderen*). The suffix -*er* can be said to produce a secondary stem used in compounds.

Another set of delimitation issues concerns the nature of the affix as a phonologically dependent, morphological unit. Affixes contrast on the one hand with clitics, and on the other with stems used in compounding. Clitics are phonologically dependent particles, i.e. particles that must attach to a host word much like an affix, whose distribution is, however, determined by syntactic rules rather than morphological rules. An example is the Italian clitic pronoun -*ti* ('you') in *incontrarti* ('to meet you'). While these clitics are generally not considered as affixes, Turkish

case and number endings as in (5) are more of a borderline case.

> (5) Ankara ve İzmire gideceğim
> 'Ankara and Izmir I go',
> i.e. I go to Ankara and Izmir

In (5), *Ankara* does not have any case ending, whereas *İzmir* is followed by the dative ending *-e*. Nevertheless, the dative ending applies to the coordinated noun phrase *Ankara ve İzmir*. Therefore, it can be argued that the dative marker is not an inflectional affix but a clitic. Similar observations can be made about the genitive marker in *the queen of England's hat*.

Compounding usually combines items that occur as independent words, e.g. *bookshelf*. Words such as *philanthropic* and *anthropomorphic* share many properties with compounds, but their constituent parts do not occur as independent words (in English). They are often called *neoclassical compounds*. In fact, *anthropo* is the Ancient Greek word for human being, a sense it also has in English words. Although it is not a word in English, it does not behave like an affix either. It has a stem-like meaning and can appear in different positions in a compound. Other elements of the same general type, e.g. *pseudo*, occur almost exclusively as a prefix, often with a reduced meaning more typical of an affix than of a stem, e.g. *pseudocultured*. It is very difficult to draw a clear borderline between stems and affixes among such neoclassical elements. The term *affixoid* is sometimes used to refer to items for which it is difficult to determine whether they are stems or affixes. An example from outside neoclassical word formation is German *reich* ('rich') in *ertragreich* ('productive, profitable'). In English, *-ful* in *successful* is similar.

Affix or Process?

One of the central questions in morphology is whether affixation should be seen as a rule applying to two elements, a base and an affix, or as a process applying to a base. Charles Hockett (1954) calls the first position Item & Arrangement, and the second Item & Process. Although all morphological processes can be described in either framework, some descriptions are more natural than others.

In a strict Item & Arrangement position, such as defended by Rochelle Lieber (1992), stems and affixes have separate entries in the mental lexicon. The suffix *-ness* is described as a noun that requires an adjective to its left. Given an appropriate concept of headedness (cf. below), this explains why *fondness* is a noun. The Item & Arrangement approach works particularly well as long as the form of the resulting word is the concatenation of the forms of the stem and the affix, as in *fondness*. Special provisions have to be made for cases where phonological processes interact with the concatenation, as in (6–8).

> (6) a. active   activity
> b. intend   intention
> (7) a. a house   to house asylum seekers
> b. to read   a good read
> (8) a. live   life
> b. extend   extent

In (6), suffixation triggers a phonological change of the base, stress shift in (6a), and a change of the final consonant in (6b). In an Item & Arrangement account, one might say that these phonological changes are triggered by the affix. In (7), there is no affix, but the relationship between the nouns and verbs is very similar to what we find in cases of genuine affixation such as *encase* and *entertainment*. In an Item & Arrangement account, one might either say that (7) is not affixation but something else (e.g. conversion), or that (7) involves zero affixes, i.e. affixes that do not have a phonological realization. The examples in (8) show that phonological changes can occur without an affix. This can be interpreted as phonological rules triggered by a zero affix or by conversion. The general approach in Item & Arrangement is that concatenation of base and affix is taken as central. Cases such as (6–8) are treated as exceptions to be fitted in.

In an Item & Process approach, as argued for by Stephen Anderson (1992), the process of affixation takes priority over the affix. In (7), the word formation process changes the syntactic properties of the input without affecting its form. In (8), the process modifies the form of the input by changing the quality of the final consonant and, in (8a), of the vowel. In (6), the process affects the form of the input by changing the stress position or the voicedness of the final consonant and appending some further phonological material, the suffix. Affixation of the type illustrated by *fondness*, the prototypical case in Item & Arrangement, is a special case in Item & Process where there is a perfect match between the input and a part of the output of the affixation process.

It is difficult to find conclusive arguments for either Item & Arrangement or Item & Process approaches to affixation. Proponents of Item & Arrangement usually claim that their approach is theoretically more restrictive than Item & Process. Proponents of Item & Process typically argue that processes such as infixation, reduplication, vowel change, consonant change, and conversion are too frequent, especially in non-European languages, to treat them as exceptional compared to 'pure' affixation. Linguistic data alone cannot determine as to which of the two approaches is correct, because every Item & Arrangement account can be reformulated into Item & Process and vice versa.

Headedness

The concept of *head of a word* only makes sense in an Item & Arrangement-based account of morphology. The head determines the syntactic category of the resulting word as well as other properties such as gender in languages that distinguish them. Thus, *fondness* is a noun because *-ness* is the head, and French *activité* ('activity') is a feminine noun because *-ité* is the head.

Different methods have been proposed to determine the head. Anna Maria Di Sciullo and Edwin Williams (1987) propose that the head is always the rightmost element. This concurs with the observation that suffixes often change the syntactic category of their base, but prefixes, as in (1b), usually do not. Exceptions include cases such as *enrich* and *debug*, where the prefixes make verbs out of an adjective and a noun. An alternative proposed by Lieber (1990) is that the last affix to attach is the head. This generalization makes *re-* the head in (1b) and accounts correctly for *enrich* and *debug*, but has problems with prefixes such as *counter-* in (9).

(9)   a. intuitive   counterintuitive
      b. example   counterexample
      c. sign   countersign

In (9), *counter-* is the only (hence last) affix attached, but the syntactic category depends on the word it attaches to. The Spanish diminutive, illustrated in (10), is problematic for both.

(10) a. el pintor   'the painter'
         el pintorsito   'the little painter'
      b. la cancyón   'the song'
         la cancyonsita   'the little song'

The final *-o* and *-a* in (10) are not a part of the affix, but so-called *word markers*, whose form is determined in this case by the grammatical gender of the word (*-o* for masculine, *-a* for feminine). Although the suffix *-sit* is the rightmost element and the only affix, it is the base that determines the gender as indicated by the article. In an Item & Process approach, there is no need for a head, because its function is subsumed in the affixation process.

References

Anderson, Stephen R. 1992. A-morphous morphology. Cambridge: Cambridge University Press.

Aronoff, Mark H. 1994. Morphology by itself: stems and inflectional classes. Cambridge, MA: MIT Press.

Booij, Geert, Christian Lehmann, and Joachim Mugdan (eds.) 2000. Morphologie—Morphology: Ein Internationales Handbuch zur Flexion und Wortbildung—An international handbook on inflection and word-formation, Vol. 1. Berlin: Walter de Gruyter.

Di Sciullo, Anna Maria, and Edwin Williams. 1987. On the definition of word. Cambridge, MA: MIT Press.

ten Hacken, Pius. 1994. Defining morphology: a principled approach to determining the boundaries of compounding, derivation, and inflection. Hildesheim: Olms.

Harris, James W. 1991. The exponence of gender in Spanish. Linguistic Inquiry 22. 27–62.

Hockett, Charles F. 1954. Two models of grammatical description. Word 10. 210–31.

Lieber, Rochelle. 1990. On the organization of the lexicon. New York: Garland.

Lieber, Rochelle. 1992. Deconstructing morphology: word formation in syntactic theory. Chicago: University of Chicago Press.

PIUS TEN HACKEN

See also **Compounding; Inflection and Derivation; Word**

African American Vernacular English

African American Vernacular English (AAVE) has been known by various names in linguistics, including Negro Nonstandard English, Black English Vernacular, and more recently Ebonics. Except for the last, each name corresponds to a particular time in American history when it was more fashionable to refer to African Americans as Negroes or Blacks, among other names.

The terms have been used to designate nonstandard English varieties spoken by less educated African Americans other than those in coastal Georgia and South Carolina, who speak Gullah, a creole. The average speakers of both vernaculars do not have specific names other than English for the ways they speak; scholars have coined the names to distinguish their

ethnolects (ethnic dialects) from standard English and from European-American vernaculars.

Ebonics, which was first used in print in the work of Robert Williams (1975), is probably the most confusing of the terms. It subsumes Gullah and thus corresponds to what may be identified broadly as *African American English*. However, in the way Williams defines it, it also refers to other language varieties of the black diaspora, including English and non-English creoles of the Atlantic (e.g. Jamaican and Haitian creoles and Nigerian Pidgin English). The name becomes quite useless when it is also extended to cover Niger-Congo languages. The average AAVE speaker might be puzzled as to why their English is lumped with a multitude of other languages they hardly understand, if at all, and why the fact that it is first of all an American linguistic phenomenon is virtually overlooked.

On the basis of literary sources, the predecessors of AAVE must have become distinct from other American English vernaculars already in the early eighteenth century (Brasch 1981). Literary representations of the time are similar to those about the Caribbean, where larger plantations with African majority populations were already booming. It is not clear whether the literary representations (mostly by British, rather than American, authors) reflect the speech of African Americans of the time or, instead, generalized stereotypes that were based on what was developing in the Caribbean. Newspaper advertisements about runaway slaves reveal that slaves born in the colonies (a large proportion by the early eighteenth century) or who had lived there for a long time did not speak differently from the European indentured servants with whom they interacted on a regular basis, before the institutionalization of segregation in tobacco and cotton plantation ex-colonies in the late nineteenth century. This history suggests that, unlike Gullah, AAVE could not have developed as a distinct variety before the late nineteenth century. Aside from letter-writing evidence, more reliable literary representations of AAVE probably also date from the nineteenth century, when most of the authors were also American born.

AAVE has become central to research in American quantitative sociolinguistics and historical dialectology since the 1960s. William Labov, William Stewart, and J.L. Dillard, together with Walt Wolfram and Ralph Fasold, focused on subsets of the following questions: (1) Is AAVE a rule-governed vernacular, and is it a full-fledged language? (2) Is it an American English dialect or is it, instead, an erstwhile creole (similar to those of the Caribbean), and therefore a separate language, which has acquired English-like features by 'decreolization'? (Putatively, this evolution would be caused by approximations by African Americans of standard English structural features, supposedly because of the mass media, mass education, and socioeconomic mobility since the abolition of slavery.) (3) How different is AAVE's system from those of standard English and European American vernaculars such as Appalachian and Ozark English? (4) How does variation work in its system and what is its significance from an evolutionary point of view?

Overall, a very large proportion of the research has been devoted to a small subset of structural features that makes AAVE peculiar, at least by its statistical distribution. This includes the following:

1. AAVE allows the loss of consonant clusters at the end of words, and this means that the final consonant is often dropped in words such as *desk, passed*, and *old* (pronounced *des', pass',* and *ol'*).
2. Diphthongs (double vowel sounds) such as [aw] and [ai], as in *south side,* become long monophthongs (simple vowels): [a:].
3. The 'r' is dropped in front of consonants and at the word end, as in *Lord* and *car* (pronounced *Lo'd, ca'*); the 'r' is pronounced before vowels, as in the words *arrive* and *road.*
4. The copula 'be' is often absent in sentences such as *She cute* and *He gon come.*
5. *Be* is used to denote invariant or repeated states of affairs, as in *He be tired/reading every time I visit him.*
6. AAVE uses multiple negation, as in *He ain go nowhere (nohow).*
7. The negative auxiliary (e.g. *didn't*) is moved to the front of the sentence, as in *Didn't nobody tell me 'hush yoh mouth',* 'Nobody told me, 'hush your mouth'.'
8. AAVE differs from standard English in various aspects of time reference, as in *I been there,* 'I have been there', and *They BEEN married* (with emphasis on *been*), 'They have been married for a long time and still are', and *He done lef when I come,* 'He had already left when I came'.
9. Plural is expressed with *dem* (in alternation with the presence or absence of the standard English plural marker *-s*), as in *dem boys*, and *an dem* signals an 'associative plural', as in *Yolanda an dem,* 'Yolanda and her friends/relatives'.
10. Many words and word meanings are peculiar to African American speech, as in *You are bad* (with *bad* pronounced emphatically and with an extra-long vowel), 'you are impressive/something else', and *you look clean,* 'you look sharp(ly dressed)'.

Interest is now growing in describing not only isolated features but also integrated subsystems of the

vernacular, highlighting, on the one hand, similarities with and differences from other American English varieties and, on the other, how subsystems such as time reference or the noun phrase are internally organized. This is a corollary of the fact that today, studies of AAVE's structure are no longer almost exclusively in the quantitative sociolinguistics paradigm, but also in theoretical linguistics.

There has also been research in African American discourse differences, especially on discourse genres, the structure and contents of some narratives, the form of language used in them, and whether or not the meanings of particular utterances are transparent. Particularly outstanding in this subarea are *playing the dozens*, in which male participants exchange witty, fictional insults directed especially at female relatives and *tall tales* or *toasts*, rhyming epics in which street culture and its heroes are celebrated in a street verbal style.

References

Brasch, Walter M. 1981. Black English and the mass media. Lanham, MD: University Press of America.

Dillard, J. L. 1972. Black English: its history and usage in the United States. New York: Vintage Books.

Labov, William. 1972. Language in the inner city: studies in the Black English vernacular. Philadelphia: University of Pennsylvania Press.

Mufwene, Salikoko S., et al. (ed.) 1998. African-American English: structure, history, and use. London: Routledge.

Poplack, Shana (ed.) 1999. The English history of African American English. Malden, MA: Blackwell.

Poplack, Shana, and Sali Tagliamonte. 2001. African American English in the diaspora: tense and aspect. Oxford: Blackwell.

Rickford, John R. 1999. African American vernacular English. Malden, MA: Blackwell.

Rickford, John Russell, and Russell John Rickford. 2000. Spoken soul: the story of Black English. New York: John Wiley & Sons, Inc.

Smitherman, Geneva. 2000. Talkin that talk: language, culture and education in African America. London and New York: Routledge.

Williams, Robert L. (ed.) 1975. Ebonics: the true language of Black folks. St. Louis, MO: Robert L. Williams & Associates, Inc.

SALIKOKO MUFWENE

See also **Caribbean; Ethnicity and Language; Gullah**

Afroasiatic

The Afroasiatic (formerly called Hamito-Semitic, also Semito-Hamitic, more recently Afrasian) phylum unites five or six language families that are clearly related, but the relations among the families are far from clear. The confusion—and the profusion of names—result from the history of the study of these languages. Known from antiquity, of course, were three of the principal Semitic languages: Hebrew, Arabic, and Aramaic in its Christian (Syriac) and Jewish literary forms. The Ethiopic literary language Ge'ez became known to European scholars during the Renaissance and was immediately recognized as a close relative of the other three. Beginning in the mid-eighteenth century, a number of ancient languages that survived only in inscriptions were deciphered, starting with Palmyrene, a close relative of Syriac, and Phoenician, which is close to Hebrew. In 1781, the name 'Semitic' was suggested for these languages as a family, on the basis of the 'Table of Nations' in Genesis 10:21–4, where Asshur, Aram, and Eber are among the descendants of Shem.

The nineteenth century brought two supremely important decipherments, those of Egyptian (in the early 1820s) and Akkadian (in its two main varieties, Assyrian and Babylonian, in the late 1840s). Certain peculiarities of the Semitic language structure, when noticed at an early stage in the decipherment of Akkadian by Edward Hincks, proved to be very helpful in making further progress, and there was soon no doubt that another major Semitic language had been discovered. With Egyptian, although Jean-François Champollion already suggested a Hebrew comparison in his initial announcement of the decipherment, and Hincks used Egyptian loanwords in the Hebrew Bible in his analyses of Egyptian, the resemblances were less obvious, and the actual kinship was suggested by the pioneering Egyptologist and linguist Richard Lepsius by 1880.

Also during the mid-nineteenth century, explorers, notably Lepsius, traveling through Abyssinia (modern Ethiopia, biblical Cush) gathered information on the languages they encountered there. Some (notably Amharic, the language of the ruling dynasty) proved to be Semitic languages, others (notably Oromo, which was then called by the insulting name Galla, and

Somali) resembled each other fairly closely and Semitic only remotely; this group was soon dubbed Cushitic. Only in the past generation has it been suggested that 'West Cushitic' is actually a separate branch of Afroasiatic (renamed Omotic), perhaps very distant from Cushitic; this was widely accepted for some time, but opinion may be shifting back to the earlier view. (The principal Omotic language is Wolaytta of Ethiopia, with some 2 million speakers, but very little is known about most of its relatives.) As Africa was partitioned among various colonial powers, naturally the investigation of Africa's indigenous languages was carried out largely by scholars from the respective metropolitan nations. Thus, much of the initial work on Cushitic was done by German and Italian linguists, while French investigators provided most of the descriptions of Berber tongues from North Africa (Morocco to Libya).

The last of the five main branches of Afroasiatic to receive attention is called Chadic, from its location near Lake Chad; the principal language of Nigeria, Hausa, is its main representative, and what little investigation was done was largely carried out by British linguists. The large number of Chadic-speaking communities, their small populations, and their remote locations, mostly in northern Nigeria, are factors that have made Chadic the least known of these groups overall.

Returning to the 'Table of Nations', we find that Ham is listed as the father of Cush and Egypt, so not unreasonably these languages that were clearly, although distantly, related to Semitic came to be called 'Hamitic'. But this proved to be a suboptimal choice for two reasons: linguistically, it implies that all the other families form a unit as against Semitic (implying that the Semitic family was the first to branch off from the others), which is not the case; and ethnographically or anthropologically, it caused the classification of languages to become enmeshed with the spurious notion of a 'Hamitic race' of dark-skinned but European-featured Africans. (In fact, of course, there is no necessary connection between language family relationship and 'race' or ethnic group.) For this reason, the name 'Afroasiatic' is now widely used for the phylum, since it simply names the two continents where it is found. This can be contracted to the convenient 'Afrasian'.

What, then, *are* the relationships among the branches of Afroasiatic? The very nature of the data makes it difficult to achieve consensus. It is as if scholars of Indo-European were faced with only Gothic and the other Germanic languages, Greek, modern Romance and Celtic, Armenian, and the modern Indic languages—analogous, respectively, to Akkadian and the other Semitic languages (an ancient tongue not ances-

tral to any of the modern ones), Egyptian (a single language attested over thousands of years), Cushitic and Omotic (whose common ancestry is disputed), Berber (a close-knit group of dialects with some ancient materials), and Chadic (a vastly differentiated family with a widely used, politically dominant variety). (The Indo-European analogy can also remind us that we have no way of knowing how many languages or entire branches of Afroasiatic have vanished without a trace—what 'Avestan' or 'Umbrian', what 'Tajik' or 'Albanian' there might have been.) Thus, it is no wonder that there are almost as many views of Afroasiatic relationships as there are Afroasiaticists. One view divides the northern groups—Berber, Egyptian, Semitic—from the southern ones—Chadic, Cushitic, Omotic. Another groups Semitic and Cushitic, plus Berber, against Egyptian and Chadic, plus Omotic.

Some characteristics that are widely shared among Afroasiatic languages are as follows. Consonants tend to be fairly numerous and vowels very few. Stops and sibilants appear in three series, voiced (*b d g, z*), voiceless (*p t k, s*), and 'emphatic', customarily marked with a dot (*ṭ, ṣ*). The 'emphasis' appears differently in different languages: in Semitic, it is mostly represented by velarization or pharyngealization (where constriction at the back of the mouth colors the consonant and adjacent vowels), but elsewhere it appears as glottalized or even as implosive varieties of these sounds. There are also a fairly large number of consonants made in the back of the mouth (including the notorious Arabic 'ayn /ʕ/). Another peculiarity is a tendency to laterals (sounds produced with air flowing past the sides of the tongue) and interdentals (with the tongue tip between the front teeth as in English *th* [θ]). Conversely, Afroasiatic languages tend to have the near-minimal vowel set /a i u/ (both long and short).

The best evidence that the Afroasiatic languages are related (rather than reflecting much borrowing of vocabulary and grammatical features) is the shared inflectional paradigms: the personal prefixes and suffixes self-evidently have a common origin throughout. The elements used to derive one verb from another also appear throughout the phylum (Table 1). It is noteworthy that both the earliest and the latest comparative grammars of Semitic, those of Heinrich Zimmern (1898) and Burkhart Kienast (2001), systematically include Afroasiatic comparisons. (Meanwhile, Carl Brockelmann, who compiled the definitive collection of comparative grammatical information on Semitic as then known (1908–1913), apologizes in the preface for excluding 'Hamitic' materials, as the data were inadequate at the time. Other writers on Semitic have followed his lead.) The well-known 'triconsonantal root' structure of the Semitic languages, however, does not appear to be

TABLE 1 Some Afroasiatic Comparisons

	Common Semitic	Ancient Egyptian	Tuareg (Berber)	Afar (Cushitic)	Hausa (Chadic)
Prefix-conjugation markers (Hausa: aspect pronouns)					
1 sing.	ʼa-	—	ø-…-ǵ	ø	náa
2 m. sing.	ta-	—	t-…-d	t-	káa
2 f. sing.	ta-…-ī	—	t-…-d	t-	kín
3 m. sing	ya-	—	y-	y-	yáa
3 f. sing	ta-	—	t-	t-	táa
1 pl.	na-	—	n-	n-	mún
2 m. pl.	ta-…-ū	—	t-…-m	t-…-n	kún
2 f. pl.	ta-…-ā	—	t-…-mt	t-…-n	kún
3 m. pl.	ya-…-ū	—	-n	y-…-n	sún
3 f. pl.	ya-…-ā	—	-nt	y-…-n	sún
Suffix-conjugation or pronominal affixes					
1 sing.	-kū	-wỉ	-i	ø	níi
2 m. sing.	-tā	-k	-k	-t	kái
2 f. sing.	-tī	-č	-m	-t	kée
3 m. sing.	ø	-f	-s	ø	-šíi
3 f. sing.	-at	-ś	-s	-t	ʼítá
1 pl.	-nū	-n	-näǵ	-tVn	múu
2 m. pl.	-tunū	-čn	-wän	-tVn	kúu
2 f. pl.	-tinā	-čn	-kmät	-tVn	kúu
3 m. pl.	-ū	-śn	-sän	-Vn	súu
3 f. pl.	-ā	-śn	-snät	-Vn	súu
Verb derivation markers					
causative	š	—	s	s	—
passive	n	—	t	m	—
reciprocal	t	—	m	t	—

shared throughout Afroasiatic; elsewhere in the phylum, 'roots' tend to have just two consonants.

The only attempts at description of the grammar of the hypothetical ancestral language 'Proto-Afroasiatic' are the two small volumes by I. M. Diakonoff (1965, 1988), and he changed his mind considerably between the two. A summary, quite close to the earlier version, may be found in his article in the *Encyclopædia Britannica* (1974). Diakonoff must be considered quite brave for even attempting to reconstruct a proto-language, given the nature of the available materials: contrast the situation in Indo-European (the real situation, not the truncated version offered above as a parallel to that in Afroasiatic). Proto-Indo-European can be reconstructed in considerable detail (and with considerable agreement among scholars) because the hypothetical community that would have spoken a language similar to what is reconstructed thrived only a couple of thousand years or so before the earliest available evidence (Hittite, Vedic, Avestan, Mycenean). We can even verify the accuracy of comparative and reconstructive methodology over such a time span because we have the precious case of the Romance languages—reconstructed Proto-Romance is pleasingly similar to Vulgar Latin as it is known

from the first centuries AD. Proto-Semitic may be considered quite reliable for the same reasons—the large amount of data and the small time depth. Contrast the situation in Afroasiatic: one branch (Semitic) well-known, one branch (Egyptian) known in considerable detail over a long time—but with no vowels known, and one branch (Berber) with a scattering of meager inscriptions from Roman times. Thus, four of the six branches are known only or primarily from modern data—and the time depth of Chadic alone is estimated as about the same as that of Indo-European.

It is no wonder, then, that among the three presentations of comparative Afroasiatic vocabulary—those of Marcel Cohen (1947; omitting Chadic, with an occasional Hausa comparison), Christopher Ehret (1995; omitting Berber), and Vladimir Orel and Olga Stolbova (1995)—there is precious little agreement. (Nor has the preliminary requirement of comparative dictionaries of each of the constituent families been accomplished, even for Semitic.) Part of the difficulty is the shortness of words in the languages, with their roots consisting of just two consonants. Thus, it is all the more difficult to proceed to the other main consideration regarding an ancestral language: the homeland of its speakers.

There are three ways of determining where a language family may have originated. One, which is hypothetical, suggests that the area with the greatest concentration of diverse, related languages is likely to be where the family originated. (It is not likely that several related language groups would all happen to move to the same locality!) For Afroasiatic, this points to somewhere in northeast Africa, near where Chadic and Cushitic (and Omotic?) are found, say Ethiopia/Sudan. Further support for this suggestion might be found in the evidence that the Egyptians entered the Nile Valley from the south and settled gradually northward. There is simply nothing but speculation as to why and whence successive populations of Semitic-speakers appeared on the scene in southwest Asia. A second technique is to examine the vocabulary that can be traced back to the protolanguage. A dishearteningly limited number of words appear to have survived from so long ago, but they seem to point to a preagricultural, nonmaritime environment. The third source of information is the accidents of discovery owing to archaeology. Terrain, climate, and politics conspire to make expeditions unlikely; but there is reason to suppose that an Afroasiatic homeland is to be sought in regions that are now the eastern Sahara Desert, but were hospitably temperate over 10,000 years ago. An archaeological horizon in the western Arabian peninsula has been suggested as coordinating with Proto-Semitic speakers, but in the absence of written records, no assignment of languages to archaeological sites can be considered certain.

What of wider connections? From time to time, Afroasiatic is claimed to be part of a still older, vastly extended superstock called 'Nostratic', which is said to incorporate Indo-European, Uralic, Altaic, and various other phyla of Asia. Any perceptible reflection of a Proto-Nostratic must be uninterpretably dim; but a question that seems not to be addressed by Nostraticists is the location of the Proto-Nostratic-speaking community—a location from which all the postulated homelands must be reachable. From this point of view, the suggestion that Afroasiatic is not a member of Nostratic, but is coordinate with the languages that make up 'Eurasiatic', is slightly more tenable. An independent suggestion, made by archaeologists who investigated Indo-European origins, that 'Afroasiatic'—from which they seem to be aware only of Egyptian and Semitic—originated on the southern coast of the Black Sea, is ipso facto untenable.

The careful study of Afroasiatic languages goes back to antiquity. The study of the Afroasiatic phylum is nearly two centuries old. The tasks that remain for researchers are largely the same as those faced by specialists in other languages: intensive fieldwork, especially in Chadic, Cushitic, and Omotic, to record ample information about these languages, and then to bring to bear all the analytic techniques of linguistic theory and comparative method. A difference with Afroasiatic is that it also includes some of the oldest written records on earth, providing a unique cross-section of human linguistic behavior across 5,000 years.

References

Brockelmann, Carl. 1908–1913. Grundriss der vergleichenden Grammatik der semitischen Sprachen. Berlin: Reuther & Reichard.

Cohen, Marcel. 1947. Essai comparatif sur le vocabulaire et la phonétique du chamito-sémitique. Paris: H. Champion.

Diakonoff, I. M. 1965. Semito-Hamitic languages. Moscow: Nauka.

Diakonoff, I. M. 1988. Afrasian languages. Moscow: Nauka.

Ehret, Christopher. 1995. Reconstructing Proto-Afroasiatic (Proto-Afrasian): vowels, tone, consonants, and vocabulary. Berkeley and Los Angeles: University of California Press.

Kienast, Burkhart. 2001. Historische semitische Sprachwissenschaft. Wiesbaden: Harrassowitz.

Orel, Vladimir, and Olga Stolbova. 1995. Hamito-Semitic etymological dictionary: materials for a reconstruction. Leiden: Brill.

Zimmern, Heinrich. 1898. Vergleichende Grammatik der semitischen Sprachen. Berlin: Reuther & Reichard.

PETER T. DANIELS

See also **Ancient Egyptian; Hausa and Chadic Languages; Semitic Languages; Tuareg and Berber Languages**

Age and Language

Comprehension and production impose many simultaneous demands on the reader to process information on a number of levels, including syntactic, semantic, and pragmatic. It has been generally assumed that these simultaneous processing demands are handled by a human's working memory and that working

memory capacity declines with age, affecting language processing. Support for this hypothesis is largely correlational. Older adults are typically found to have smaller working memory spans than young adults, and such span measures are found to correlate with language-processing measures. A particular source of difficulty for older adults is the production and comprehension of complex syntactic structures involving multiple levels of sentence embeddings that impose high processing loads on working memory. The spontaneous use of such constructions declines with advancing age, and their use contributes to the breakdown of reading and listening comprehension.

Text processing is particularly vulnerable to the effects of aging; both cross-sectional and longitudinal research shows that older adults' reading comprehension is impaired, as is their listening comprehension. Further, word and name retrieval is disrupted by aging, contributing to an increase in tip-of-the-tongue difficulty in finding words. It is unclear as to whether working memory affects the immediate processing of words and sentences, or whether working memory exerts its affects on the processes involved with the storage, recall, and recognition of linguistic information. The general slowing of cognitive processes with age may also contribute to older adults' processing problems by affecting the speed of lexical access and other component processes. Neural inhibitory mechanisms also appear to weaken with age and permit the intrusion of irrelevant thoughts, personal preoccupations, and idiosyncratic associations during language processing. These irrelevant thoughts compete for processing resources, such as working memory capacity, and impair older adults' comprehension and recall.

A controversial issue in the study of aging and language is 'off-target verbosity'. A minority of older adults have been observed not only to talk a great deal but also to drift from topic to topic, weaving into their conversations many unrelated and irrelevant topics. Off-target verbosity appears to be related to poor performance on tests of the function of particular brain regions, as well as to psychosocial stress, extraversion, limited social support, and decreased social interaction. Whereas off-target verbosity has been cited as providing strong support for inhibitiory deficit theory, others have argued that this speech style is limited to social settings in which older adults construe their task differently than do young adults—as a monolog, responsive to an internal chain of associations.

The debates over working memory capacity limitations and inhibitory deficits are central to the general study of gerontology. These core research issues have been supplemented by two emerging topics: research examining the effects of Alzheimer's disease on language production and comprehension, and research examining how to improve older adults' comprehension of language. Task demands, lapses in attention, response biases, or other cognitive deficits contribute to the syntactic comprehension problems of adults with dementia.

Semantic and lexical processes are particularly disrupted by Alzheimer's dementia. Many researchers have found that tests of implicit word knowledge reveal few differences between dementing adults and normal older adults, and lexical knowledge is commonly found to be preserved. Other researchers have concluded that the structure of semantic knowledge is destroyed by Alzheimer's dementia. As a result, performance on verbal-fluency tasks, such as generating examples of categories or words beginning with a specific letter, is impaired, picture and object naming is hindered, and word associations are destroyed. The semantic network of adults with Alzheimer's dementia may be intact but inaccessible, perhaps due to the general, task-independent slowing of cognitive processes, including lexical access. General slowing may increase progressively with disease severity. A breakdown of inhibitions may also contribute to many of the word retrieval problems experienced by adults with Alzheimer's disease if they are unable to inhibit inappropriate word associations.

A special speech register, sometimes termed 'elderspeak', has been described as an accommodation in communicating with older adults, especially those with dementia. This accommodation is characterized as involving a simplified speech register with exaggerated pitch and intonation, simplified grammar, limited vocabulary, and a slow rate of delivery. Many of the characteristics of elderspeak, such as its slow rate, exaggerated prosody, and simplified syntax and vocabulary, resemble the characteristics of other speech registers such as 'motherese', or the speech used by adults when communicating with small children. Elderspeak may be evoked by the actual communicative needs of the elderly individual as well as by (culturally based) negative stereotypes of older adults, and it is often viewed by older adults as insulting and patronizing. Older adults, especially those in nursing homes, may adapt to situational demands by becoming more accepting of elderspeak. Some form of elderspeak may enhance communication with older adults, especially those with Alzheimer's dementia, but little systematic research has been undertaken to assess intervention and training programs.

In sum, the study of language and aging has focused on the effects of processing limitations on older adults' comprehension and production. Nonetheless, it is important not to lose sight of three general points. First, many aspects of language processing do not change with age, particularly lexical access and semantic

memory, but are vulnerable to the ravages of Alzheimer's disease. Other aspects of language, particularly syntax, may be more resistant to Alzheimer's disease, yet susceptible to age-related decline due to working memory limitations or inhibitory deficits. Second, there has been a gradual shift toward the use of sophisticated methodologies such as the study of self-paced reading times, eye movement patterns, and brain activation patterns to study how aging and dementia affect language processing. Third, the shift toward developing and evaluating practical applications, as seen in the growing body of research on elderspeak, continues. The linkage of basic research on language processing with everyday practicalities has led to the development of consumer standards and guidelines for presenting medical information, and for electronics and telecommunications.

References

Arbuckle, Tannis, and Deloris Pushkar Gold. 1993. Aging, inhibition, and verbosity. Journal of Gerontology: Psychological Sciences 48. 225–32.

Caplan, David, and Gloria Waters. 1999. Verbal working memory and sentence comprehension. Behavioral and Brain Sciences 22. 114–26.

Charness, Neil, Denise C. Parks, and Barbara A. Sabel (eds.) 2000. Communication, technology, and aging. Doylestown, Pennsylvania: Springer.

Hasher, Lynn, and Rose T. Zacks. 1988. Working memory, comprehension, and aging: a review and a new view. The psychology of learning and motivation, Vol. 22, ed. by Gordon H. Bower. New York. Academic Press.

Kemper, Susan. 1992. Language and aging. The handbook of aging and cognition, ed. by Fergus I. M. Craik and Timothy A. Salthouse. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kemper, Susan, and Tamara Harden. 1999. Disentangling what is beneficial about elderspeak from what is not. Psychology and Aging 14. 656–70.

Kemper, Susan, and Reinhold Kliegl (eds.) 1999. Constraints on language: aging, grammar, and memory. Norwell, MA: Kluwer Academic.

Light, Leah L., and Deborah M. Burke (eds.) 1988. Language, memory, and aging. New York: Cambridge University Press.

Park, Denise C., Roger W. Morrell, and Kim Shifren (eds.) 1999. Processing of medical information in aging patients: cognitive and human factors perspectives. Mahwah, NJ: Lawrence Erlbaum Associates.

Ryan, Ellen B., Howard Giles, Giampiero Bartolucci, and Karen Henwood. 1986. Psycholinguistic and social psychological components of communication by and with the elderly. Language and Communication 6. 144–66.

Snowdon, David A., Susan J. Kemper, James A. Mortimer, Lydia. H. Greiner, David. R. Wekstein, and William R. Markesbery. 1986. Cognitive ability in early life and cognitive function and Alzheimer's disease in late life: findings from the Nun study. Journal of the American Medical Association 275. 103–13.

Wingfield, Arthur, and Elizabeth A. L. Stine-Morrow. 2000. Language and speech. The Handbook of Aging and Cognition, 2nd edition, ed. by Fergus I. M. Craik and Timothy A. Salthouse. Mahwah, NJ: Lawrence Erlbaum Associates.

SUSAN KEMPER

See also **Aphasia; Psycholinguistics; Working Memory**

Agreement

Agreement is a means by which languages signal the presence of a grammatical relation—usually the subject of a sentence, but sometimes also the object and/or the indirect object. Typically, it manifests itself as a prefix or a suffix on the verb, as with the English third-person singular marker ***-s*** (e.g. *John **lives** in Maryland*). Some languages have so-called rich agreement—that is, separate forms for nearly every person/number combination. French represents such a case:

(1) *Rich agreement (French)*

1SG	je parl-e	'I speak'
2SG	tu parl-es	'You speak'
3SG	il parl-e	'He speaks'
1PL	nous parl-ons	'We speak'
2PL	vous parl-ez	'You speak'
3PL	ils parl-ent	'They speak'

The features most often associated with agreement are person, number, and gender. Given the logistics of speech, there are at least three persons: those referring to the speaker (first), the hearer (second), and everything else (third). Some languages have a fourth person—obviative—reserved for entities further removed from the speaker/hearer's point of view. Inuktitut (Eskimo) and Algonquian are known for this.

Number usually breaks down into singular and plural, but 'dual' represents another possibility. English recognizes two numbers in its pronominal system, but

as far as agreement is concerned, only third-person subjects are marked: cf. *He understands* but *We understand*.

In addition to person and number, agreement may also involve the gender features of a subject and/or objects. French nouns are divided into two grammatical genders—masculine and feminine. When noun phrases headed by these nouns appear in subject position, agreement can be discerned on predicate adjectives and particles (M = masculine, F = feminine):

(2)   *Gender agreement (French)*

    a. Le soleil est monté   'The sun (M) has risen'

    b. La lune est montée   'The moon (F) has risen'

                   [PARTICIPLE]

    c. Le soleil est brillant   'The sun (M) is bright'

    d. La lune est brillante   'The moon (F) is bright'

                   [ADJECTIVE]

Feminine forms are marked with an extra vowel that—although not pronounced today—were presumably audible at an earlier stage of the language. Of course, grammatical gender is not to be confused with biological gender; the division of nouns into gender classes should be considered as an aid to reference and/or acquisition (German has three genders, for example).

Agreement and Case

As a grammatical category, agreement appears to function like case-markers—special morphemes that attach to noun-phrases (NPs) (rather than verbs) that signal their grammatical function within the sentence (e.g. as subject, object, etc.). In Japanese, for instance, NPs marked with [-ga] are interpreted as subjects, while those marked with [-o] are understood as objects (TOP = topic; COMP = complementizer):

(3)   *Case-marking and grammatical relations (Japanese)*

    Ichiro-wa   Taro-**ga**   hirugohan-**o**
    Ichiro-TOP [Taro-SUBJ   lunch-OBJ

    tabeta-to   iita
    ate]-COMP   said
    'Ichiro said that Taro ate lunch'

In traditional grammar, the case of the subject is called nominative, while that of the direct object is accusative. Since only subjects are associated with agreement in English, this may be taken as a sign of nominative case. Objects are not overtly marked with accusative case, but their grammatical function can be determined by their position following the verb.

Another way in which agreement appears functionally similar to case is through its interaction with transitivity. In languages like Japanese, transitive and intransitive subjects take the same case-marker, and in opposition to objects. This is known as a nominative/accusative case-marking system. English is essentially the same with respect to agreement—that is, it pertains to both types of subject. In some languages, however, intransitive subjects and transitive objects pattern in the same manner, in contrast to transitive subjects. Dyirbal (Australian) has a case-marking system of this type, known as ergative/absolutive (transitive subjects are marked with ergative case). Languages that employ agreement, rather than case to signal grammatical functions can exhibit the same pattern, as in Mam (Mayan) below (ABS = absolutive; ASP = aspect; ERG = ergative):

(4)   *Ergative/absolutive agreement (Mam)*

    a. ma   chiin-x-a
       ASP 1SG/ABS-go
       'I went'
                  [INTRANSITIVE]

    b. ma   chin-ok   t-tzeeq'an-a
       ASP 1SG/ABS   2/ERG-hit
       'You hit me!'
                  [TRANSITIVE]

Why is it important to indicate grammatical functions? Many linguists believe that languages use these devices to identify the participants in the event specified by the verb. In a sentence like 'The President hires young assistants', for example, the verb calls for two participants: someone who initiates the action (the President), and someone/something affected by it (young assistants). The first is sometimes called an Actor, the second (in this sentence) a Patient or Undergoer. Grammatical functions signaled by subject agreement and accusative case (the latter by virtue of its position following the verb) ensure that each participant is interpreted correctly, or that the sentence does not mean 'Young assistants hire the President'. Languages that employ neither case-marking nor agreement must obviously find some other way of interpreting such potentially ambiguous sentences.

Agreement and Inflection

Agreement is often called an inflectional category, as opposed to lexical ones like nouns, verbs, and adjectives. Other inflectional categories include tense, mood, and aspect. Sometimes, agreement is fused with one or more of these to form a single entity. This is the case with the English suffix -s: not only does it signal a third-person singular subject but also present tense and indicative mood. Agreement disappears in the past tense and/or indicative mood: cf. *John **lives** in Maryland now, but last year he **lived** in Maine; Mary wants to stay, but it is necessary that she **leave*** (NOT: ***leaves***). Still, there is some debate as to whether agreement is truly absent from these forms or simply covert—that is, 'present but invisible'. In German, for example, agreement is perfectly compatible with the past tense, and in Palauan (Austronesian) with the irrealis mood (similar to nonindicative):

(5) *Agreement and past tense (German)*

1SG	ich glaub-t-e	'I believed'
2SG	du glaub-t-est	'You believed'
1SG	wir glaub-t-en	'We believed'

(6) *Agreement and irrealis mood (Palauan)*

	REALIS	IRREALIS
1SG	ak-	ku-
2SG	ke-	omo-
3SG	ng-	le-

Object agreement is often associated with the inflectional category aspect. Again, in Palauan, objects of completed events trigger 'perfective' agreement on the verb, whereas objects of ongoing ones do not (IMPF = imperfective; NM = noun-marker; P-preposition; PF = perfective; R = realis):

(7) *Agreement based on aspect (Palauan)*

a. ng-chillebed-ii a bilis
   R3SG-hit-PF3SG NM dog
   'S/he hit the dog'
   [PERFECTIVE]

b. ng-milengelebed er a bilis
   R3SG-hit (IMPF) P NM dog
   'S\he was hitting the dog'
   [IMPERFECTIVE]

The relationship between agreement—which basically pertains to nouns—and tense, mood, and aspect (a property of verbs) is therefore sometimes quite intricate, and poses a major challenge for linguists attempting to explain it.

Nonstandard Cases

While subjects normally trigger agreement in English-tensed clauses, there are a number of cases in which it is suppressed and/or the status of the subject itself is called into question. One case involves 'subjunctive' mood, following verbs like *suggest* and *demand*: cf. *I demanded that he **be** on time; We quickly suggested that she **leave*** (compare: *…that she **should** leave*, which is also acceptable but does not show signs of agreement). Other languages show reduced forms of agreement under negation, or when one grammatical function or another is questioned:

(8) *Reduction of agreement (Chamorro/ Austronesian)*

a. Ha-fa'gasi si Juan i kareta
   3SG-wash Juan car
   'Juan washed the car'
   [NORMAL AGREEMENT]

b. *Hayi ha-fa'gasi i kareta?
   who 3SG-wash car
   'Who washed the car?'
   [WH-QUESTION/NORMAL AGR]

c. Hayi fuma'gasi i kareta?
   who wash+UM car
   'Who washed the car?'
   [WH-QUESTION/SPECIAL AGR]

The existential *there* construction in English arguably has two subjects: one occurring in clause-initial position (filled by *there*), the other post-verbally and signaling agreement: cf. *There's a man standing on the corner* vs. *There **are** three men standing on the corner.* Nevertheless, speakers sometimes take the clause-initial subject as the agreeing one, resulting in sentences like *There's four things that can be said about him.* Other nonstandard (but occurring) breakdowns involve plural complements of singular subjects (*Neither one of you guys **are** going to win*), and coordinate (hence plural) subjects with singular marking on the verb: *The fact that we were bombed, and that the markets reopened so quickly suggests that the foundations of our country are quite strong.*

Exotic Forms of Agreement

Agreement in natural language can deviate significantly from single forms that encode features of person and number. In Potawatomi (Algonquian), each of

these features can be associated with a separate morpheme (OBV = obviative):

(9) *Split agreement (Potawatomi)*

1SG	n-wapm-a	'I see him'
2SG	k-wapm-a	'You see him'
3SG	w-wapm-a-n	'He sees him (OBV)'
1PL	n-wapm-a-mun	'We see them/him (OBV)'
2PL	k-wapm-a-wa	'You see him'
3PL	w-wapm-a-wa-n	'They see them/him (OBV)'

Here, the prefixes [n-], [k-], and [w-] represent first, second-, and third-person subjects, regardless of their number. The suffixes [-wa] and [-mun] indicate plurality (the latter also encodes features of the object, i.e. plurality and/or obviation). Interestingly, the same prefixes and suffixes also appear on nouns to indicate possession and number:

(10) *Nominal agreement*

(Potawatomi; c = alveopalatal affricate)

1SG	n-ciman	'my canoe'
2SG	k-ciman	'your canoe'
3SG	w-ciman	'his/her canoe'
1PL	n-cimanwa-nan	'our canoe'
2PL	k-ciman-wa	'your canoe'
3PL	w-ciman-wa	'their canoe'

In the larger picture then, agreement must be regarded as more than just a property of verbs. 'Subjects' of NPs (possessors) corrolate with sentential subjects: it is no accident that the possessive marker [-s] in English has the shape it does (cf. *John's dog*). In Irish and other Celtic languages, prepositions inflect for the person/number features of their objects; this grammatical relation roughly correlates with sentential objects.

To fully appreciate how complex agreement can be, consider that objects of transitive verbs in Potawatomi can also be associated with separate morphemes of person and number. In (9) above, the suffix [-a] signals that the object is a third person, while [-n] registers its plurality and/or obviation; also notice how the subject plural marker [-wa] intervenes between these two suffixes.

Inuktitut (Eskimo) takes a different tack in the distribution of agreement features, using single forms called portmanteau to encode the person and number of both subject and direct object (ABS = absolutive; AL = allative; CMT = comitative; PRT = participle; REL = relative):

(11) *Portmanteau (Inuktitut)*

a. tuni-jara   nutarar-mut
give-PRT/1SG/3SG child-AL

arna-up   taku-jaa-nut
woman-REL   see-PRT/3SG-AL
'I gave it to the child that the woman saw'

b. arna-up   pilauti-milk
woman-REL knife-CMT

angut   tuni-jaa
man (ABS)   give-PRT/3SG/3SG
'The woman gave the man the knife'

In (11a), the form [-jara] signals a first-person subject and third-person object, while in (11b) [-jaa] indicates two third persons.

Evolution of Agreement

One might well ask why agreement is so pervasive in language, or where it comes from. Many morphemes appear to originate as independent pronouns that slowly undergo a process of phonological reduction and subsequent attachment. A comparison of pronouns and agreement from Udi (Caucasian) demonstrates this most clearly:

(12) *Pronouns and agreement (Udi)*

	AGREEMENT	PRONOUNS
1SG	-zu, -z	zu
2SG	-nu, -n, -ru, -lu	un
3SG	-ne, -le, -re, -n	meno, kano, seno
1PL	-yan	yan
2PL	-nan, -ran, -lan	van, efan
3PL	-q'un	met'ovon, kat'ovon, set'ovon

Although the match-up is far from complete, there is enough similarity of form to allow for a theory of language change. Independent/emphatic pronouns may pass through an intermediate stage of cliticization on their way to becoming agreement morphemes. Pronominal clitics also share the fundamental features of person, number, and gender. French has a series of subject-markers traditionally regarded as pronominal: *je, tu, il,* etc. Nevertheless, these forms cannot be separated from the (auxiliary) verb like their English counterparts; cf. **Il vraiment est retourné* vs. *He really has gone home.* This suggests that *je, tu,* and *il* are clitics, and in the frequent presence of emphatic pronouns may

someday evolve into a new series of agreement morphemes (cf. *Moi, je vais retourné*). Also, note that (except for first- and second-person plurals) the traditional agreement suffixes of modern French are no longer pronounced.

Theory of Agreement

How do linguists analyze agreement? Usually, a local relationship is assumed to hold between the element that triggers it (e.g. the subject NP), and an inflectional head, which may be comprised of tense, mood, or aspect. Agreement is the visible sign of this relationship. This explains why only verbs in tensed clauses show agreement—at least in English. In structural terms, the subject NP is called the specifier of the clause, attached to but outside the immediate domain of inflection: [NP [verb+inflection, etc.]]. This also parallels the internal structure of the NP, where articles like *a, the,* and *some* are attached to but outside the immediate domain of the noun: [the [dog]]. In fact, possessed NPs (e.g. *John's dog*) can be assigned essentially the same structure as a clause if the possessor (*John*) is regarded as a specifier, and the sign of possession ([-*'s*]) as a kind of inflection. It is probably no accident that English uses the same sound to express both specifier–head relationships.

The function of agreement—that is, as a means of highlighting one participant in the event described by the verb—can be explained in the following way. First, a verb's meaning is more closely comprised of its non-Actor participants than its Actor participants. The verb *to kill,* for instance, can involve essential information that is separate from the person who initiates the action (cf. *kill an hour, kill a motion, kill the mood, kill the light*—in addition to the standard meaning 'cause to die'). Languages typically recognize this distinction through the mechanism of case-marking, and non-Actor participants are often made visible through cases selected by the verb. Actors, on the other hand, are less likely to be singled out by case. Instead, they must rely on some other mechanism to be rendered visible, such as agreement. Inflection then—the head of every clause—provides a specifier position that can accommodate one non-case-marked participant for every verb.

References

Anderson, Steven. 1992. A-morphous morphology. New York: Cambridge University Press.

Aronoff, Mark. 1994. Morphology by itself: stems and inflectional classes. Cambridge, MA: MIT Press.

Comrie, Bernard. 1989. Language universals and linguistic typology, 2nd edition. Chicago: University of Chicago Press.

Croft, William. 1990. Typology and universals. New York: Cambridge University Press.

Givon, Talmy. 1984. Syntax: a functional–typological introduction. Amsterdam: John Benjamins Press.

Marantz, Alec. 1992. On the nature of grammatical relations. Cambridge, MA: MIT Press.

Pinker, Stephen. 1984. Language learnability and language development. Cambridge, MA: Harvard University Press.

MARK CAMPANA

See also **Austronesian; English; French Language; German; Grammar, Traditional; Japanese**

Ainu

The Ainu are an indigenous group found mainly on the northern Japanese island of Hokkaido. Physically and linguistically distinct from the neighboring Japanese, at one time the territorial range of the Ainu extended throughout the northern Japanese island of Hokkaido (Yezo), including the Kurile islands to the northeast and southern Sakhalin (Karafuto) to the north. Evidence from place names, however, suggests that at one time the Ainu had also settled in northern Tohoku, on the main Japanese island of Honshu, and on Russia's Kamchatka Peninsula as well. Due to resettlement of the Sakhalin Ainu (1875) and the Kurile Ainu (1884) today, however, the majority of the remaining Ainu are situated on the island of Hokkaido. It should be noted that the Ainu language found in northern Japan is unrelated to the Turkic language bearing the same name that can be found in western China.

The genetic and linguistic history of the Ainu has been the topic of much debate. Physical features such as increased body hair (when compared to the Japanese) and more distinct facial features make the Ainu appear Caucasian to many people and as such the Ainu have sometimes been considered Caucasian (Caucasoid) in ancestry. Other scholars have sought to draw a relationship between the Ainu and native

American groups. Recent studies conducted with DNA testing, however, suggest the possibility that the Ainu are Mongoloid in origin.

Numerous attempts have been made to classify Ainu linguistically, including theories that place Ainu in the Malayo-Polynesian, Paleo-Asiatic, and Ural-Altaic language families. At present, there are several competing theories on the relationship of Ainu with other languages. The first and most widely offered explanation is that Ainu is a language isolate, with at present no apparent relationship with other languages. The second theory rests on the assumption that Ainu is related to the Altaic language family, and languages such as Japanese and Korean. Some studies arguing the latter position suggest that if Ainu is at all connected to other Altaic languages, then the splitting of Ainu from Japanese and other Altaic languages must have occurred at least 10,000 years ago. However, Alexander Vovin (1993), in his work on reconstruction of Proto-Ainu (the hypothesized ancestor language), suggests a connection to the Austro-Asiatic languages of the south. Further complicating matters has been the historical lack of a writing system, and as a consequence, a written record of Ainu over time. The relatively rapid transition from a viable, functioning language with monolingual native speakers to one that is, at present, nearly extinct has further confounded attempts to conclusively place Ainu among the world's languages.

The Ainu have had a continuous presence in Hokkaido for at least the past 1300 years. The first record of the Ainu on Yezo occurs in the Japanese *Kojiki* (A.D. 720) calling them 'Emishi', or those outside of the Japanese law. The Ainu refer to themselves as *Utari* or 'compatriot' in Ainu, and this term is in more frequent use at present because of the sometimes negative connotations that have been leveled on the term 'Ainu' by the Japanese. The word *Ainu* simply means 'human'.

While no written form of Ainu existed until recently (Ainu is now written in Romanized form or with the Japanese *katakana* script), the Ainu culture has developed a rich oral tradition most well known for their epic poems called Yukar. Yukar follow traditional patterns in their subject matter with some of the Yukar told from the perspective of an animal, others based on activities of ancestors, and still others are hero tales. The Yukar differ from the spoken language, tending to exhibit more complex (polysynthetic) word structures and more conservative features than spoken Ainu. These stories have served as a rich source of linguistic material as they maintain archaic forms, providing a rich source of data attesting to earlier forms of the Ainu language. Yukar has come to be a collective term

for all forms of epic story in Ainu but as Masayoshi Shibatani notes,

> In the strict sense of the term yukar refers only to the heroic verse, mythic epics being more specifically referred to as kamuy yukar, mat yukar, or oyna. There are as well prose-style old stories and folktales.

As noted above, Ainu is a polysynthetic language, where words tend to be composed of multiple morphemes. This is particularly the case in classical Ainu. In contrast, colloquially spoken Ainu tends to be more analytical, with more of the morphemes appearing as independent words, suggesting the possibility of a long-term change in progress.

In terms of its grammar, Ainu is a S(ubject)–O(bject)–V(erb) language. But, as seen in the following example, word order is important in determining the grammatical relationship in Ainu.

Kamuy aynu rayke.
bear person kill
'The bear killed the man.'

Aynu kamuy rayke.
person bear kill
'The man killed the bear.'

Phonologically, Ainu has a five-vowel system (/a/, /i/, /e/, /o/, /u/) with syllable-initial vowels preceded by a glottal stop. The consonants of Ainu, 12 in all, include the stops of /p/, /t/, /k/, and the glottal stop, sibilant /s/, glottal fricative /h/, affricate /c/, semi-vowels /w/,/y/, nasals /m/ and /n/, and the liquid /r/. Ainu has a pitch accent system in which syllables are pronounced with high or low pitch. Its syllable structure is C(onsonant)V(owel) or CVC. The exception to this, however, is the Sakhalin dialect, which contained long and short vowel contrasts, also allowing for CVV to occur. Vowel sequences tend to be avoided in other forms of Ainu. Vocalic euphony or vowel harmony, the assimilation of a vowel to another vowel in the word or phrase, is one of the more notable aspects of the sound system of Ainu.

The Ainu vocabulary is rich with references to nature, reflecting the close relationship held between the people and natural phenomena. Most Ainu villages (called *kotan*) were located alongside rivers and coasts and former Ainu settlements are easily recognizable in place names today with endings such as -*nai* or -*pet* (-betsu is the Japanese representation), both meaning 'river', as found in the place names of Wakkanai, Noboribetsu, and Monbetsu. The importance of salmon and bears to Ainu culture is represented by the

abundance of lexical items related to these two animals. Particularly, the role of the bear was an important part of Ainu culture, and they were often kept in villages for ceremonial events and as food.

Within Ainu, three distinct dialect groups are generally recognized, the Kurile, Sakhalin, and Hokkaido dialects, each representing the distinct regions where the Ainu once inhabited. The last speaker of the Sakhalin dialect is believed to have died in 1994. Within these dialect groups, further distinctions can be made, such as that between southern and eastern Hokkaido Ainu. At least 19 dialects of Ainu have been identified, most of which are now extinct.

It is estimated that the number of Ainu descendants (including those of mixed Ainu-Japanese heritage) number around 24,000, with those of solely Ainu descent being very small. Postwar census figures classify the Ainu as Japanese for census purposes; therefore, only estimates of the number of remaining Ainu in Hokkaido are possible. Discrimination aimed at the Ainu by Japanese as well as a Japanese desire to view Japan as a single homogeneous society have further caused many Ainu to conceal their heritage. It is believed that the actual number of Ainu in Hokkaido may be double or triple the figure of the government estimate of 24,000 (1995). The Ainu face regular discrimination in employment and marriage opportunities. Such discrimination and fear of disclosing their Ainu ethnicity have, over time, contributed significantly to the loss of the Ainu language.

The Ainu language and culture have been impacted dramatically by the influx of the Japanese into Hokkaido, which began as early as the 1300s. The peak of the Ainu culture occurred in the thirteenth and fourteenth centuries, prior to the large-scale arrival of the Japanese (wajin), which began in the fifteenth century. From this point onward, battles with the Japanese in 1457, 1669, and 1789, as well as diseases such as smallpox, measles, cholera, and tuberculosis all took their toll on the Ainu.

In 1550, an agreement was reached between the Ainu and the Japanese to allow a Japanese settlement area on Yezo. This would be a turning point in Ainu-Japanese relations. Over the next three centuries, Japanese control of Hokkaido increased and finally in 1869 the official government-sanctioned colonization of Hokkaido began. The Japanese population of Hokkaido quickly grew to over 1 million, placing the Ainu in a minority. From this point onward, the Ainu were systematically stripped of their cultural and linguistic identity, land, and hunting rights through a forced assimilation policy of the Japanese government. Also during this period, disease again struck the Ainu, leading to a decrease in their population.

By 1871, many Ainu customs, including the tattooing of women's mouths and the burning of a familial household upon the death of a family member, were banned. The Ainu language was also banned as the Japanese government set out to assimilate the Ainu into becoming Japanese and officially classifying them as 'former aborigines'. In 1899, the Hokkaido Aborigine Protection Act was created; yet, despite its name, the act did little more than increase the efforts to assimilate the Ainu. The Japanese government encouraged the Ainu to shift from their traditional role as hunter/gatherers to farming. At the same time, Ainu lands were ceased. When land was redistributed to the Ainu, they were given the least fertile lands. In 1901, Ainu children were forced to attend Japanese primary schools and Ainu children began learning Japanese as their first language. The assimilationist policies of the Japanese government resulted in a near elimination of the Ainu language and culture by the mid-twentieth century. Today, there are no monolingual native speakers left.

Today, the Ainu are fluent Japanese speakers. In spite of the past Japanese efforts to eliminate the Ainu language, bilingual speakers of Ainu and Japanese do exist, however, and recently modest efforts have begun to continue the teaching of the Ainu language and traditions to those willing to learn. Ainu language use also continues in ceremonial and ritualistic use and in events performed for tourists. A preserved Ainu village at Shiraoi is maintained for educational and tourism purposes.

In 1946, the Ainu Association of Hokkaido was founded, seeking to bring increased attention to the fight for Ainu rights, the preservation of the language and culture, and a reevaluation of the Hokkaido Aborigine Protection Act.

In 1994, Shigeru Kayano became the first Ainu member of the Japanese Diet (parliament). Subsequently, in 1997, the Japanese government finally and officially recognized the Ainu as the indigenous people of Hokkaido and created the Ainu rights law. This law was designed to replace the Hokkaido Former Aborigine Protection Act. It requires local governments in traditional Ainu areas to support promotion of the Ainu culture, including promotion of the Ainu language. As a result, government-subsidized weekly Ainu language radio broadcasts are offered in Hokkaido with free texts that are widely available. There are also 13 state-supported language classrooms in traditional Ainu villages throughout Hokkaido that seek to provide opportunities for students to learn Ainu. A bi-weekly Ainu language newsletter has also been produced.

The 1997 Ainu rights law provided official recognition to the Ainu and a degree of legimaticy to the

efforts aimed at preservation and revitalization of the language. Whether the Ainu language will undergo a successful revitalization is dependent upon the ability of the Ainu to increase awareness among the Japanese of what might be lost, should the Ainu language and culture disappear.

References

Batchelor, John. 1889. An Ainu–English–Japanese dictionary. Tokyo: Y. Kumata; Iwanami-Syoten, 4th edition, 1938.

Brochure on the Ainu. Sapporo, Japan: The Ainu Association of Hokkaido, n.d.

Chiri, Mashiho. 1952. Ainugo ni okeru boin chouwa. Hokkaido Daigaku Bungaku Kiyou 1. 103–18.

Hattori, Shiro. 1959. Nihongo no Keitou. Tokyo: Iwanami.

Kayano, Shigeru. 1994. Our land was a forest: an Ainu memoir. Boulder, CO: Westview Press.

Maher, John, and Kyoko Yashiro. 1995. Multilingual Japan: an introduction. Journal of Multilingual and Multicultural Development 16(1). 1–17.

Patrie, James. 1982. The genetic relationship of the Ainu language. Oceanic Linguistics Special Publication No. 17. Honolulu: University of Hawai'i Press.

Shibatani, Masayoshi. 1990. The languages of Japan. Cambridge: Cambridge University Press.

Vovin, Alexander. 1993. A reconstruction of Proto-Ainu. Leiden: E. J. Brill.

SHAWN M. CLANKIE

See also **Altaic; Japanese**

Akan and Nyo Languages

Nyo, like Left Bank, belongs to the (New) Kwa subfamily of the Niger–Congo group. Nyo has more languages than the Left Bank. There are five daughter languages in Nyo: Agnéby, Attié, Avikam–Alladian, Potou–Tano (which accounts for over half of New Kwa in terms of both the speakers and geographical area), and Ga–Dangme. Potou–Tano is composed up of six language groups: Ega, Potou, Tano (previously called Volta–Comoé), Lelemi (and its related languages), Logba, and Basila–Adele. Westermann classified Avikam–Alladian, Agnéby, Attié, and Potou under the Lagoon subfamily; and Lelemi (and its related languages), Logba, and Basila Adele as Togo–Remnant languages.

Agnéby: The Agnéby languages, a Central Nyo group, are spoken in the Ivory Coast in the Agnéby and Bandama river basins by 230,000 people. Adioukrou, Abiji, and Abbey are the three languages in this group. Adioukrou (Adjukru) is spoken in the north of the Ebrié Lagoon westward of the Agnéby river by about 70,000 people. Abidji is spoken by about 80,000 people in the Abidjan District and in the subdistricts of Sikensi and Dabou. Abidji has two main dialects: Enyembe and Ogbru. Speakers of Abiji tend to be bilingual in Adioukrou, Baoule, and Jula. Important linguistic features include tone and nasalization on syllables. Abbey, spoken by about 80,000 people, is spoken around Agboville and to the north of the Abiji and Krobou areas. Examples of words and their plurals taken from Adioukrou are as follows:

libr 'red'   - ebr 'reds'
lufu 'white' - efu 'whites'
lorɲn 'snake' - morŋn 'snakes'

Attié: Attié is spoken in Côte d'Ivoire by about half a million people living in Abidjan, Anyama, Alep, Adzope, Affery, Agou, Akoupe, and Yakasse–Attobrou. The three main dialects of Attié are Anaindin (Nindin), Ketin, and Bodin (the most prestigious and populous). Most Attié speakers are bilingual in Abbey, Anyin, Baoule, Ebrie, and Jula.

Avikam–Alladian: It is a Western Nyo language spoken in the Western Ivory Coast. It has two languages: Avikam and Alladian. Avikam is spoken in the basin of the Bandama River by 12,000 speakers, whereas Alladian is spoken by about 15,000 people on the east of Avikam.

Potou–Tano: The six language groups identified under Potou–Tano are Ega, Potou, Tano, Lelemi (and its related languages), Logba, and Basila–Adele.

Ega: Ega was originally classified as Kru, but Richards (1982) proved that it is Kwa. It is located inside the Kru area. Like other Kwa languages, Ega has a reflex of the cognate (words maintained by languages of common ancestry) nyɔ /ɲõ/ 'two'.

Potou: It is spoken in the Eastern Ivory Coast Lagoon area by about 75,000 people. Potou has two languages:

Ebiré (spoken by 60,000 people in the north of the Ebiré Lagoon from the Agnéby River to the Potou Lagoon in the east) and Mbatto (spoken by about 15,000 between the Potou Lagoon and the Comoé River).

Tano: Tano, previously called Volta–Comoé, has more speakers than any of the subfamilies of the Potou–Tano group. Tano has four subfamilies: Krobou, Western Tano (Abouré and Eotilé), Central Tano (Bia, Akan), and Guang (Southern and Northern). Krobou has about 6000 speakers and is spoken within the Agnéby language area in the Ivory Coast. Abouré and Eotilé are also spoken in the Ivory Coast. Abouré is spoken by about 40,000 people in the Comoé River basin near the sea, whereas Eotilé is spoken by 3000 people in Vitre #I and Vitre #II on the Ebiré Lagoon to the north of Grand Bassam.

Bia: The Bia languages, sometimes referred to as Nzema–Anyi–Baule, are spoken by nearly two and a half million people in the Bia, Tano, Nzi, Bandama, and Comoé River basins and along the south coast of Ghana and the Ivory Coast. The two main subgroups are: Northern Bia (comprising Anyi, Baule, Chakosi (Anufo)); and Southern Bia (Nzema and Ahanta). Anyi and Nzema are spoken in Ghana and in the Ivory Coast, Baule in the Ivory Coast, Ahanta in Ghana, and Chakosi in Ghana and Togo. Nzema has two dialects: Nzema and Evalue. Anyi also has two dialect groups: Anyi (Aowin) and Sehwi (Sanvi). Anyi has other dialects like Indene, Moronou, Bona, Djoablin, Ano, and Bini (all in the Ivory Coast), and Afema and Brossa in Ghana. Baule, with nearly one and a half million speakers, is the largest language group in the Bia subfamily. Chakosi is found in an enclave in the Gur territory on both sides of the Ghana–Togo border. In Togo, Chakosi is found in Sansanne–Mango and in Ghana, in the Chereponi area. Most speakers of Nzema (about 120,000) are found in Ghana; in the Ivory Coast, there are about 50,000 speakers.

Important linguistic features in the Bia group include consonant mutation (in which consonants change some of their phonetic qualities in certain phonetic or grammatical contexts. For example, in Nzema /k/ changes to /h/ between vowels. /k/ in *kɔ* 'go' becomes /h/ in *jihɔ* 'she's gone') and vowel harmony (cooccurrence restrictions in the distribution of vowels, i.e. the vowels are grouped into two sets and in words of more than one syllable; only vowels from one set occur. For example, in Nzema, words *otosu* mean 'he still cries' and *ɔtɔkʊ* 'he still fights'; all the vowels in *otosu* are produced with an advanced tongue root position, whereas those in *ɔtɔkʊ* are produced with a retracted tongue root). Other linguistic features include double articulated sounds

(single sounds with two different places of articulation, e.g. /kp, tp, gb/ . For example, in Chakosi, the words *kpiɛ* 'old man' and *gblaki* 'to faint' have /kp/ and /gb/, respectively). Bia languages also have tone terracing, where pitch is lowered toward the word end. Plural is formed through affixation. For example, in Baule, the plural marker {-mũ} can be added to any count noun. For example, *wa* 'child' *wamũ* 'children,' ti 'head' *timũ* 'heads.' There are also independent nasalized vowels (phonemic nasalization of vowels). For example, in Baule, *wu* means 'see', whereas *wũ* means 'die'.

Akan

Akan is the largest of the Central Tano group of languages. Until the 1950s, these clusters of mutually intelligible varieties did not have a single name. Some of the dialects (such as Fante, Bono, and Wasa) were viewed as different languages. However, a unified Akan orthography was designed by the Akan Orthography Committee in the 1950s and since then, the name Akan has been adopted as the name of the language.

Akan is spoken by over ten million people in Ghana and in the Ivory Coast. In Ghana, Akan is spoken between the rivers Tano and Volta, stretching from the coast to the inland. Abron (Bono) or Dormaa is spoken in Ghana and in Côte d'Ivoire by about a million people. In Côte d'Ivoire, Akan (Abron) is spoken in the northwest by about 60,000 people. Speakers of Abron speak and understand Asante Twi. Wasa is spoken by nearly 300,000 people in Southwestern Ghana. It has two main dialects called Amenfi and Fianse. Wasa and Abron are mutually intelligible.

In Ghana, Akan is spoken by nearly half of the population as their mother tongue and by two-thirds of the population as a second language. Akan has 11 dialects: Fante (Mfantse), Akuapem, Asante, Agona, Dankyira, Asen, Akyem, Kwawu, Ahafo, Bono (Abron), and Wasa. The Akuapem, Fante, and Asante dialects have different officially recognized orthographies and are used on radio and TV. Quite recently, a unified orthography for all the Akan languages has been adopted.

The Twi (Asante–Akyem–Kwawu) dialects are the largest dialects, with a total population of over seven million. Fante, spoken along the coast, has over two million speakers and four important dialects: Agona, Gomua, Abura, and Anomabu. Akuapem is spoken on the Akwapim Hills between the Akyem and Ga areas.

Important linguistic features include a Subject–Verb–Object word order, example:

S   V   O
Kofi   *nom*   *nkwan* 'Kofi drinks soup'
Ama   *dɔ*   *Kwame* 'Ama loves Kwame'

Akan languages also have two types of vowel harmony: cross-height and rounding. With respect to cross-height harmony, we observe that the vowels are grouped into two sets: advanced /i, e, æ, o, u/ and unadvanced /ɪ, ɛ, a, ɔ, ʊ/. In any word of two or more syllables, only vowels of one harmonic set occur. For example, *obewu,* 'She'll die', has vowels from the advanced set, whereas *ɔbɛwʊ,* 'She'll give birth', has vowels from the unadvanced set of vowels. With respect to rounding harmony, in some of the Akan languages, especially in Fante, vowels of verbal affixes agree with those of a verb root by being advanced or unadvanced, rounded or unrounded. For example, *muruwu,* 'I'm dying', has only rounded advanced vowels, whereas in *miridzi,* 'I'm eating', all the vowels are both advanced and unrounded. Rounding harmony is also found between vowels of nominal roots and affixes in the Asante and Akyem dialects. For example, the vowel of the noun suffix /o/ in *owuo* 'death' agrees in rounding with the vowel /u/ of the noun root, *owu* 'die'. The vowel /e/ suffix of the noun root *esie* 'hill' harmonizes with those of the root /e, i/ by being unrounded.

The Akan languages have uncommon sounds like the alveolo-palatal semivowel /ɥ/, labio palatalized alveolo-palatal affricate /tɕɥ/, occurring in words like we /ɥi/'chew' and Twi /tɕɥi:/ 'name of some Akan dialects'.

Another important feature of the Akan languages is tone terracing (especially downdrifting, where, during the production of long utterances, there is a gradual drop in pitch heights if high tones are interspersed with low tones. The drop in pitch could be such that an utterance final high tone could be lower in pitch than an utterance initial low tone. The final drop in pitch could also reach the bottom of a speaker's pitch range. For example, in: *Pàpá Kòfí dìdí wìé kòtìé* 'Father/Mr. Kofi finishes eating and goes to listen', the initial low tone on /pà/ of *Pàpá* is higher in pitch than the final high tone on /é/ of *kòtìé.*

Guang: The Guang subfamily, which is spoken by about half a million people in Ghana, The Ivory Coast, Togo, and Benin, has more languages than the Central Comoé group. However, the languages in this subfamily are for the most part spoken by small numbers of people. This group of languages subclassify into: North, North-East, Central, and South. North Guang languages include Gonja, Nkonya, Nawuri, and Choroba. Gonja (Ngbanito) is spoken in a large area from Bole to Kamabuie near Salago and between the White Volta, Volta, and Daka rivers in Ghana. In the Ivory Coast, Gonja is spoken in the Bondoukou and in Togo it is spoken at Semere and is called Bazantcé. The Gonja variety spoken in the Republic

of Benin is referred to as Chombulon or Tshummbuli (similar to a Ghanaian Central Guang variety called Nchumburu).

Southern Guang subdivides into Coastal Guang (a dialect cluster—Awutu, Efutu, and Senya). Awutu is spoken in Winneba, Senya Breku, Obutu, and Bawjiase in Ghana, spoken by about 200,000 people, and Hill Guang is spoken by about 150,000 people on the Akwapim Hills and on the left bank of the Volta Lake. Important dialects include Cherepong, Larteh, and Anum (Gwa).

Other Guang subfamilies include Lelemi-Lefana (Buem), Akpafu (Siwu), and Santrokofi spoken in Ghana and Togo by about 45,000 people; Logba spoken by about 4,000 people on the Ghana–Togo border; and Basila–Adele, comprising the Basila and Adele dialects, is spoken by about 10,000 in both Ghana and Togo. Basila has three subdialects: Gokolodjya, Gilempla, and Giseda. Adele has more speakers in Togo (7,000) than in Ghana (3,000).

Like the Akan group of languages, the Guang languages have advanced tongue root vowel harmony. For example, in the following Efutu words, all the vowels are either advanced *oɲi* 'man'or unadvanced *ɔkʊ́ɔ̀* 'again'. The harmony may operate over entire sentences. For example, *muudi nu* 'She is eating (food)'; *mʊ́ǹwɔ́* 'She did not go'.

Negation is formed through affixation and consonant alternation. For example, in Siwu, in negating commands, voiceless stops change into voiced stops and voiced stops become nasals. Also, the morphemic particle daa is preposed to the main verb. Example, *kɛ̀lɛ̀* 'go' dàágɛ́lɛ̀ 'don't go'; *kpɛ̀* 'weed'; dàágbɛ́ 'don't weed'.

Ga–Dangme: Ga–Dangme has two languages: Ga and Dangme. The Ga–Dangme subfamily is spoken in Ghana on the southeastern coast around Accra and inland. Ga is spoken in Ghana along the coast in and around Accra (Ghana's capital), and in the Densu and Kpone basins by about 500,000 speakers. Dangme (Adangme) is spoken along the coast (to the east of Ga) and at the mouth of the Volta and also in the Akwapim Hills. Like Ga, Dangme has over half a million speakers. Dangme has six dialects: Ada, Ningo, Prampram, Shai, Krobo, and Osudoku. Krobo is the largest dialect, with about 300,000 speakers. Structurally, Ga has two tones, High and Low, and a downstepped high tone (a high tone whose pitch has been lowered. Downstep is marked with an exclamation mark placed before the syllable). For example, *ékòmé* 'one', *gbé!kɛ́* 'boy'. Plurals are formed through affixation, (suffixation). For example, *tsò* 'tree', *tsèi* 'trees'; *tʃù* house' *tʃùi* 'houses'. There are three tones, High, Mid, and Low, in Dangme.

References

Dakubu, Mary Esther Kropp. 1988. The languages of Ghana. London: Kegan Paul International.

Greenberg, Joseph. 1955. Studies in African linguistic classification. New Haven: Compass Publishing Company.

Stewart, John. 1989. Kwa. The Niger–Congo languages, ed. by John Bendor-Samuel. New York: University Press of America.

SAMUEL OBENG

See also **Niger-Congo**

Albanian

Albanian is the name English speakers use to refer to the language of Albania (Albanians call the language 'shqipe'), as well as for the branch of the Indo-European family to which it belongs. As a branch of Indo-European, Albanian is a sister to many language groups from Europe to South Asia: Germanic (including English, Swedish, and German, for example), Romance (including French, Spanish, and Italian), Slavic (including Russian, Serbian, and Polish), Indic (including Hindi, Gujarati, and Urdu), Hellenic (Greek), Armenian, and Celtic (including Irish and Scottish Gaelic). Persuasive evidence for the kinship of Albanian with other Indo-European languages was first advanced in the middle of the nineteenth century and continues to be amassed. In addition to regular correspondence of sounds between particular words and parts of words in Albanian with those in several already established Indo-European languages, the evidence lies in features of Albanian verb conjugation and noun declension that reflect a common origin with those languages: for example, the irregular forms of certain verbs (e.g. Albanian *është*=English *is*=Latin *est*=Sanskrit *asti*, while Albanian *jam*=English *am*=Latin *sum*=Sanskrit *asmi*).

Albanian linguists claim that the predecessor of modern Albanian is the ancient Illyrian language, but since little is left of Illyrian, verification for this claim is elusive. The earliest texts clearly identifiable as Albanian are scarce and go back only to the fifteenth century; we have no record of the characteristics of the language during earlier periods, although extensive documentary evidence exists for many other Indo-European languages. To complicate the task of tracing its origins, Albanian, like English, has adopted and adapted a great substance of the vocabulary from other languages throughout its history—in earlier times, basic vocabulary from Latin and Turkish and more recently, technical vocabulary from Italian, Russian, and English.

Today, Albanian is spoken by more than six million people—about three million in Albania itself, another two million in Kosovo, some 300,000 in Macedonia, and at least 700,000 more living mainly in Montenegro, Greece, Turkey, Italy, and the United States. In recent years, both voluntary and forced emigration of Albanians has greatly increased the presence of speakers of the language outside their homelands.

During its known history since the fifteenth century CE, Albanian has been associated with two main dialects: Tosk in the south and Gheg in the north. After World War II, the government made increasingly successful efforts in Albania to adopt a unified written form of the language, called Standard Literary Albanian, and in 1968 this form of Albanian was also adopted for schools and mass media by Albanians in Kosovo. However, for ordinary conversation, few Albanians in Albania and Kosovo completely abandoned their own dialects. In rural areas, 'unified Albanian' may be scarcely used and known only from television and radio, and even well-educated, urban Albanian speakers exhibit regional and social differences in their language, which appear in pronunciation, vocabulary, and grammar. The speech of Albanian speakers living outside Albania reflects their Gheg or Tosk origins.

A modern alphabet was devised for Albanian in 1908. This 36-letter alphabet includes six digraph letters (*th, dh, sh, zh, gj, nj*) and two special characters (*ë and ç*), and was adopted quite readily by speakers. For Gheg words, a carat (^) is used to mark nasalized vowels, but the vowel length so distinctive to Gheg dialects remains unmarked in written Albanian.

In addition to its genetic membership in the Indo-European language family, Albanian shares certain specific features with other Indo-European languages in the Balkan area. For example, like Romanian, Albanian marks 'definiteness' (in English, marked by the word *the*) by endings on nouns: Albanian *gjysh*, 'grandfather' becomes *gjyshi*, 'the grandfather';

similarly, Romanian *bunic*, 'grandfather' becomes *bunicul*, 'the grandfather'.

Particular features of Albanian seem particularly striking to non-Albanians. Already mentioned is the vast number of loan words from the international vocabulary—such as *heroik, burokrat, standardi-zon*—as well as many words from Turkish (during the 400-year occupation of Albania under the Ottoman Empire), such as *penxhere* (window), *fukara* (poor or pauper), and *çoban* (shepherd). Foreign names are generally spelled in Albanian to approximate their sound—to an Albanian ear—in the language from which they derive, so that Heather Lockwood would become Hedhër Lakvud.

As in many other languages, Albanian indicates the grammatical function, or case, of a noun with noun endings. Thus, *gjyshi*, 'the grandfather', is the subject of a verb, but as the direct object of a verb it takes the form *gjyshin*, and as the indirect object of the verb, it takes the form *gjyshit*.

The verb system in Albanian is quite complex. Simple verbs not only have different endings to indicate whether their subject is the person speaking ('first person'), the person spoken to ('second person'), or someone or something else ('third person'), but also whether that subject includes the person speaking ('first person plural'), the person addressed ('second person plural'), or is entirely composed of others ('third person plural'). Those endings also indicate whether the time is present or past; if past, the ending indicates whether the verb applies over time or at a particular moment.

Another set of endings indicate whether the subject is itself the object of the verb. These 'medio-passive' endings are used for a range of functions that are quite different in English. For example, *e rruan* means 'they shave him', but *rruhen* can mean, 'they are shaved' (passive), 'they shave themselves' (reflexive), or 'they shave one another' (reciprocal). Still other sets of verb endings can express the speaker's surprise or dismay, and yet others may express a wish.

Pronouns also take different forms for varying grammatical functions. For example, as subjects of verbs, objects of verbs, and objects of certain prepositions, first- and second-person pronouns have the accented forms shown in Table 1.

Albanian lacks corresponding forms for third-person pronouns, but instead uses a complex system of 'pointing' words that convey the perceived distance of the referent, as well as the referent's number, gender, and case (see Table 2).

Unaccented pronoun forms indicate that the verbs immediately following have an object (direct or indirect) (see Table 3).

These forms appear whether or not an accented object is also present in the sentence, so that 'Agim told me' is *më tha Agimi*, and 'Agim told *me*' is *më tha Agimi mua*; similarly, 'Agim told her' is *i tha Agimi*, and 'Agim told *her*' is *i tha Agimi asaj*. Unaccented third-person direct objects may occur in combination with any unaccented indirect object, yielding the forms shown in Table 4.

The position of the unaccented pronoun (or combination) in Albanian is unique among Indo-European

TABLE 1

	English equivalent	Subject	Direct object	Indirect object	Prepositional object
First person					
Singular	I, me	unë	mua	mua	meje
Plural	we, us	ne	ne	neve	nesh
Second person					
Singular	you (thou)	ti	ty	ty	teje
Plural	you all (you)	ju	ju	juve	jush

TABLE 2

	English equivalent	Subject Far	Subject Near	Direct object Far	Direct object Near	Indirect object Far	Indirect object Near	Prepositional object Far	Prepositional object Near
Third person		Far	Near	Far	Near	Far	Near	Far	Near
Singular									
Masculine	that one: he, it, him	ai	ky	atë	këtë	atij	këtij	atij	këtij
Feminine	that one: she, it, her	ajo	kjo	atë	këtë	asaj	kësaj	asaj	kësaj
Plural									
Masculine	they, them	ata	ky	ata	këta	atyre	këtyre	asosh	kësosh
Feminine	they, them	ato	kjo	ato	këto	atyre	këtyre	asosh	kësosh

TABLE 3

Person	Number	English equivalent	Direct object	Indirect object
First				
	Singular	me	*më*	*më*
	Plural	us	*na*	*na*
Second				
	Singular	you (thou)	*të*	*të*
	Plural	you all (you)	*ju*	*ju*
Third				
	Singular	him, her, it	*e*	*i*
	Plural	them, 'em	*i*	*u*

TABLE 4

Indirect object	Direct object	Combination	Example	Meaning
më	*e*	= *ma*	*ma dha Agimi*	Agim gave it to me
më	*i*	= *m'i*	*m'i dha Agimi*	Agim gave them to me
të	*e*	= *ta*	*ta dha Agimi*	Agim gave it to you
të	*i*	= *t'i*	*t'i dha Agimi*	Agim gave them to you
na	*e*	= *na e*	*na e dha Agimi*	Agim gave it to us
na	*i*	= *na i*	*na i dha Agimi*	Agim gave them to us
ju	*e*	= *jua*	*jua dha Agimi*	Agim gave it to you
ju	*i*	= *jua*	*jua dha Agimi*	Agim gave them to you
i	*e*	= *ia*	*ia dha Agimi*	Agim gave it to him/her
i	*i*	= *ia*	*ia dha Agimi*	Agim gave them to him/her
u	*e*	= *ua*	*ua dha Agimi*	Agim gave it to them
u	*i*	= *ua*	*ua dha Agimi*	Agim gave them to them

languages in one particular aspect: in positive commands, it may come directly after the verb, but if the command is addressed to a plural 'you', then it comes before the verb ending.

Finally, Albanian is unusual among languages in its use of a set of 'attributive articles', which indicate that the word after it refers to a noun or pronoun, its 'referent', which may or may not be present in the sentence. The choice of which of the four members of this set (*i, e, të, së*) is to be used is determined by the grammatical form of the referent. Pure adjectives must always be preceded by one of these articles: *këndim i mirë* 'good singing' with the adjective *i mirë* 'good' —contrast the adverb *mirë* 'well' as in *këndon mirë* 'He or she sings well'.

Any noun with a 'genitive' case ending can be preceded by one of these attributive articles, functioning much like English *of* to indicate possession—*djali i Gjonit* (the son of John, or John's son)—but more often to indicate some looser relationship in which the noun after the attributive article narrows the scope of its referent. For example, such a construction may be used to identify:

- the substance of which something is made: *buka e grurit* (the wheat bread, or literally, 'the bread of the wheat')
- a quality that characterizes the referent: *aktet e trimërisë* (acts of bravery)
- the object of an action: *shkelja e rregullit* (the violation of the rule)
- the subject of an action: *ardhja e miqve* (the arrival of friends)
- the whole of which the referent is part: *gjysma e bukës* (half of the loaf)
- the identifying element of the referent: *muaji i korrikut* (the month of July)
- a characterizing quality after the verb 'to be': *është i gojës* (he is articulate, or literally, 'is of the mouth')
- the universe of a superlative: *më i madhi i djemve* (the biggest of the boys)

Among the small national languages of Europe, Albanian can claim a special pride in the consistency of its standard orthography and grammar, and the richness of its vocabulary.

References

Camaj, Martin. 1984. Albanian grammar. Wiesbaden: Otto Harrassowitz.

Hamp, Eric. 1972. Albanian. Linguistics in Western Europe. Current Trends in Linguistics, Vol. 9, ed. by Thomas A. Sebeok. The Hague: Mouton.

Hubbard, Phillip L. 1985. The syntax of the Albanian verb complex. New York & London: Garland Publishing.

Huld, Martin E. 1984. Basic Albanian etymologies. Columbus, OH: Slavica Publishers.

Newmark, Leonard. 1957. Structural grammar of Albanian. Bloomington: Indiana University Publications of the Research Center in Anthropology, Folklore, and Linguistics.

Newmark, Leonard. 1980. Spoken Albanian. Ithaca, NY: Spoken Language Services; revised edition, 1997.

Newmark, Leonard. 1998. Oxford Albanian–English dictionary. Oxford: Oxford University Press.

Newmark, Leonard, Philip Hubbard, and Peter Prifti. 1982. Standard Albanian: a reference grammar for students. Stanford: Stanford University Press.

Newmark, Leonard, and Vladimir Dervishi. 1999. Albanian handbook with English and Albanian glossaries. Kensington, MD: Dunwoody Press.

Zymberi, Isa. 1991. Colloquial Albanian. London and New York: Routledge.

LEONARD NEWMARK

See also **Indo-European 1: Overview**

Algeria

Algeria, officially the Democratic and Popular Republic of Algeria, is the second largest country in Africa with 919,595 square miles after Sudan, stretching from the shores of the Mediterranean to the Sahara Desert. It has a population of 31,736,053 (2001 estimate) and is divided into 48 provinces called wilayat and 700 local communes. The population consists of Arabs, Berbers (Kabyles being the largest group followed by Chaouia) and people of mixed Arab and Berber ancestry. Arabs form 80–83%, Berbers 16–20%, and others 1%.

Linguistic independence came rather late to Algeria. During the Muslim expansion of the seventh and eighth centuries, the Arabic language spread from the Arabian Peninsula to North Africa, supplanting the local languages. During the Ottoman Empire, the Arabic language was suppressed, as elsewhere in the Arab World. Later, the French conquest of Algeria in 1830 further complicated the linguistic situation. The authorities introduced French as the sole official language. Their main goal was to replace the Arabic culture by a more 'civilized' culture. The use of Arabic was generally limited to a small minority of religious scholars, and to the rural regions, where Arabic religious schools were set up to teach the *Qur'an.*

After gaining its independence from France in 1962, the Algerian government declared Arabic as the official language, which was a difficult task as very few Algerians were actually literate in Arabic. Teachers were imported from Arab countries such as Egypt and Syria. And in the 1970s, the Arabization policy required that street and shop signs be in Arabic, despite the fact that 60 % of the population could not write or read Arabic. Over the years, Algeria struggled between the pressure to eliminate any legacy from the French colonialism, to meet the demands of the Kabyles to adopt Tamazight as a second official language and to teach in schools, and to bear the costs of rapid Arabization. Arabic is held in high regard because it is the language of the Qur'an. With the spread of Islam, Arabic has become the most used language among the living Semitic languages. It is related to Hebrew, Aramaic, and Amharic. Subgroups of Arabic are Classical Arabic, Eastern Arabic, Western Arabic, and Maltese. The modernized version of Classical Arabic, which is referred to as Modern Standard Arabic, is the language that Arab countries use as their official language.

The linguistic situation in Algeria is complicated, and is a matter of intense discussion within the country itself. Algeria's population diversity is reflected in the large array of languages spoken. Seventeen languages are believed to be spoken in Algeria, Arabic and, most recently, Tamazight, being the official languages. The majority of Algerians speak Arabic, followed by French and Berber. Modern Standard Arabic, a simpler version of Classical Arabic, is taught at schools and is used in formal meetings and in the media, but is not used for ordinary conversation. The Algerian dialect, known as Darja, is used in everyday life. And increasingly, the vernacular is being used in theater and in novels because it is believed to reflect the Algerian culture.

After gaining independence, Algeria proceeded to 'arabize' education in order to rid the country of French influence. However, the fact that the country was trying to rediscover its roots while forcing its population into 'modernity' has pushed it to maintain French as a language of science and technology. Thus French became associated with high social and educational status. Arabization on the other hand was pursued for ideological reasons related to nationalism and religion.

Despite the efforts to introduce Arabization to rid the country of the cultural traces left by the long French colonialism, Algerians continue to use French for formal and informal conversations. In fact, French is considered by many the 'unofficial' official language, as it is used in most formal administrative meetings, gatherings, and various other functions. Modern Standard Arabic is used only in formal education, by the media, and for written purposes.

The policy that the Algerian government has adopted toward minority language groups has resulted in tension among the different language users. Arabization, the language policy aimed at expanding the role of Modern Standard Arabic to the detriment of French and local dialects and languages, led to dissatisfaction and resentment among the Berber ethnic minority and the French educated population. Opponents of this policy argued against making Arabic the exclusive language of instruction in public schools when French remains the language of science instruction at the universities.

Most Algerians speak Algerian Arabic, with educated and clerical masses speaking Modern Standard Arabic in formal situations. Arabs usually do not

speak Berber because it is not taught in schools. However, most Berbers speak Arabic and French, especially the younger generation. Currently, there is an active movement to regain the Berber language and teach it in schools. As of recently, Tamazight is being taught in the Kabyle major cities, but not in the non-Berber areas, although there are pockets of Berbers in every major city.

Even though the Algerian census does not gather information about household languages, the number of Berber speakers has been estimated to be as high as 6,000,000 speakers. The Berbers are considered to be the original inhabitants of North Africa. They have lived in Algeria since 3000 BC, or, according to some historians, as early as 8000 BC. The word Berber comes from the Greeks, who used this term to refer to the people of North Africa. Although Berber uses an ancient script called *tifhagh*, it is primarily a spoken language. Currently, Roman script is used to write the Berber language. Most Berbers, especially the Kabyles, maintain their own language and a sense of ethnic identity. The Chaouia, however, tend to be more assimilated to Arab culture than the Kabyles. The Kabyles refer to themselves as *Imazighen*, and to their language as *Tamazight*. The linguistic division between Arabic speakers and *Tamazight* speakers stems mainly from the government policies that favor the Arabic language over the other local languages such as Tamazight. For years, the Algerian government suppressed the Berber languages, depriving those who did not speak, write, or read Arabic from basic human rights such as reading the newspaper and watching television.

The Berbers, also known as *Amazigh* in North Africa, are the original inhabitants of North Africa. Because the Berbers have mingled with other ethnic groups over the centuries, especially the Arabs, their identification is purely linguistic and not racial. The Berber language (*Tamazight*) is primarily oral. Its written form is little known and rarely used. Like Arabic, the Berber language is a branch of the Afro-Asiatic language family. Many languages derive from the Berber language family. Some of the Berber languages, the number of speakers, and the geographical areas where they are mostly spoken are given in table.

As of 2003, 41 years after independence, French is still taught as a second language and is introduced at the second-grade level. French also remains the language of science instruction in all universities after a semi-unsuccessful attempt to arabize the educational system, including higher education. Attempts to substitute English for French as the second national language in the mid-to-late 1990s also failed. However, since 2003, English is being taught at the seventh-grade level, a year earlier than in previous years.

Berber Language	Number of Speakers	Geographical Areas
Kabyle	Up to 6,000,000 (1998)	In the Djudjura mountain range, and along the northern central and eastern coastal region, east of Algiers Main cities: Tizi-Ouzou, Dellys, Bejaia
Chaouia	1,400,000 (1993)	South and southeast of the Grand Kabylie region and south of Constantine, in the Aurès Mountains Main cities: Batna, Ain el Baidha, Ain Mlila
Chenoua	15,000–75,000 (1996)	Small towns east of Algiers
Tachelhit	Unknown	Southern Algeria near the Moroccan border
Tahaggart	25,000 (1987)	Southern Algeria in the Hoggar region Major cities: Djanet and Tamanrasset
Taznatit	40,000 (1995)	Around the city of Timimoun
Tumzabit	70,000 (1995)	Mzab region, 330 miles south of Algiers Main city: Ghardaia

Because the spoken Algerian Arabic is a simplification of Modern Standard Arabic, there is less distinction regarding number and gender. For example, the dual form and the feminine plural do not exist in the Algerian dialect. The Modern Standard Arabic masculine plural form is used in Algerian Arabic for dual, feminine and masculine, and for the feminine plural. Algerian Arabic consists of three major dialects spoken in three major cities: Algiers (center), Constantine (east), and Oran (west). These major dialects differ considerably from the dialects spoken by nomadic people in the Sahara.

Algerian Arabic vs. Modern Standard Arabic

Phonological Differences

Modern Standard Arabic is written from right to left and uses a root system, usually consisting of three consonants. For example, the English word 'merge' would be made up of the root 'm-r-g'. In Arabic, the word 'islam' is made up of the root 's-l-m'. The Arabic script uses symbols for only consonants and

long vowels. The short vowels are not written and are represented by diacritics. However, although morphological and grammatical meaning is indicated by vowels, the vowel diacritics are usually omitted. The reader has to infer its particular meaning and pronunciation from the context.

Even though the Algerian dialect is a variant of Arabic, its vocabulary differs considerably from Modern Standard Arabic. Many words originate from sources other than Modern Standard Arabic, such as Turkish, Spanish, Italian, and most of all *Tamazight* (Berber) and French. However, most of the vocabulary comes from Modern Standard Arabic with phonetic, phonological, and semantic changes. The examples below show Modern Standard Arabic words with Algerian Arabic correspondences. The elements in which the Algerian Arabic words differ from the Algerian Arabic ones are indicated in bold.

Modern Standard Arabic	Algerian Arabic
Assimilation:	
taji:'u	**d**ji (she comes)
tazu:ru	**d**zu:r (she visits)
tadu:mu	**d**du:m (it lasts—fem.)
Sadr	**z**dar (chest)
ni**S**f	nu**SS** (half)
Metathesis:	
yartaɛid	ya**tr**ɛid (he shivers)
laɛana	**n**ɛal (he cursed)
Dissimilation:	
finja:n	finja:**l**
ba:dhanja:n	badhanja:**l**
Phone Substitution:	
shajara	**s**ajra (a tree)
Monophthongization:	
zawj	**zu:**dj (two)
Sayf	**Si:**f (summer)
zayt	**zi:**t (oil)
Semantic Changes:	
yudi:ru	ydi:r
(he directs)	(he does)
'ajlis	'uqɛud
(sit down—when someone is standing up)	
'uqɛud	tgaɛɛad
(sit down—when someone is lying down)	

The French Influence

The Algerian dialect contains a good number of borrowed French words. However, the borrowed vocabulary is subject to Arabic rules of grammar and pronunciation. For example, French has nasal vowels, but Arabic does not. Therefore, the nasal vowels in borrowed French words are denasalized (1). Also, French sounds that do not exist in Arabic are substituted by Arabic sounds (2). As the Algerian dialect relies heavily on emphatism, loan words are no exception (3). Arabic has three short vowels, *a*, *i*, and *u,* and three corresponding long vowels, *a:*, *i:*, and *u:*. As a result, French vowels that do not exist in Arabic are replaced by one of the three vowels (short or long) (4). French influence extends to syntax as well (5).

French	Algerian Arabic
(1) *bouchon*	*bushu:n* (cork)
(2) *il roule*	*yru:li* (he wanders) The French *r* is replaced by the Arabic *r*.
une serviette	*serbita* (towel) The sound *v* is replaced by b.
une savate	*Sabba:t* (a shoe)
(3) *il sonne*	*ySu:ni* (he rings) The French s is replaced by the emphatic S.
une place	*bla:Sa* (a place)
(4) *un bureau*	*bi:ru* (a desk, an office) The French *ü* is replaced by *i:*.
(5) *une table*	*Tabla* (a table) The Arabic feminine indicator *a* is added.

The linguistic situation in Algeria depends on the future political situation. The Algerian people believe that if the Muslim fundamentalists come to power, the trend will be toward arabization of all sectors of life; on the other hand, if democratization is successful, bilingualism will be stressed with the French language used for the technical and hard sciences, and Arabic for social sciences and civil services. After years of political unrest, the political and cultural situations are far from being resolved. While a solution is being sought for the linguistic conflict, Modern Standard Arabic remains a foreign language for some of the older Algerians, especially the Berbers. However, recent attempts to make Tamazight a second national language have been somewhat successful. In 2002, President Bouteflika announced that the Tamazight language would become an official language in Algeria alongside Arabic.

References

Benremouga, Karima. 1987. A survey of a typical eastern dialect of Algerian Arabic: phonology and morphology. Master's Thesis. Lawrence, KS: University of Kansas.

Benremouga, Karima. 1997. Spoken Algerian 'Arabic': An overview. [Ed. by In Xingzhong Li, Luis, López,

and Tom, Stroik, Papers from the 1997 Mid-America Linguistics Conference.] Columbia, MO: University of Missouri.

Grimes, Barbara F. (ed.) 1992. Ethnologue, languages of the world. Dallas, TX: Summer Institute of Linguistics.

Harik, Elsa M., and Donald G. Schilling. 1984. The political of education in colonial Algeria and Kenya. Athens, OH: Ohio University Center for International Studies.

Ruhlen, M. 1987. A guide to the world's languages. Vol. 1: classification. London: Edward Arnold.

Saad, Zohra. 1993. Language planning and policy attitudes: a case study of Arabization in Algeria. Doctoral Dissertation. New York: Columbia University.

Talmoudi, Fathi. 1984. The diglossic situation in North Africa: a study of classical Arabic/dialectal Arabic diglossia with sample text in mixed Arabic. Göteborg, Sweden: ACTA Universitas Gothoburgensis.

KARIMA BENREMOUGA

See also **Arabic; French Language**

Altaic

Altaic, in its broadest conception, is a family of languages composed of four subfamilies (Japanese-Okinawan, Mongolic, Tungusic, and Turkic) and two language isolates (Ainu and Korean). However, the status of this family remains a matter of controversy. On the one hand, some scholars who accept the genetic association of Mongolic, Tungusic, and Turkic (hereafter, Altaic proper) do not agree that there is a demonstrable relationship between Altaic proper and Japanese-Okinawan, Korean, or Ainu. On the other hand, some scholars question even the notion of Altaic proper. The skepticism often rests on purely methodological grounds; that is, although they accept the likelihood that Mongolic, Tungusic, and Turkic languages all stem from a common ancestor, some linguists doubt that enough evidence exists to prove this connection conclusively. In other instances, linguists reject even the likelihood of Altaic proper; they argue that the similarities among these language families are better explained by borrowing and contact-induced changes than by a genetic association.

Ainu is spoken on Hokkaido, the northernmost island of Japan, and on Sakhalin Island, which is part of the Russian Federation. Ainu is spoken only as a second language, primarily by older people. Only about 1% of the 18,000 to 25,000 ethnic Ainus claim an ability to speak the language, and the number of fluent speakers is probably a fraction of this small number.

Ainu, as is typical of Altaic languages, has a basic word order of subject–object–verb, but the connection with other Altaic languages is tenuous at best. Such a connection was proposed by John Street (1962) and was argued more forcefully by James Patrie (1982), but few accept Ainu as being part of the Altaic family.

Japanese-Okinawan, a subfamily of Altaic, consists of two branches. The first branch consists solely of Japanese, which is spoken by approximately 130 million people worldwide. The second branch, called Okinawan or Ryukyuan, consists of a set of ten languages, most of which are mutually intelligible to some degree. There are roughly 900,000 speakers of Okinawan languages, although very few of the languages are being transmitted to children at this point. Japanese and the Okinawan languages share roughly 70% of their vocabulary; they are not mutually intelligible.

The earliest written records of Japanese date back to the early eighth century and are composed in Chinese characters. Since then, two different syllabic writing systems derived from Chinese characters have been developed for Japanese: Hiragana and Katakana. Both are syllabic; i.e. each symbol stands for a syllable. Additionally, there is now a phonetic writing system, Romaji, which uses the Roman alphabet. Here, the symbols stand for individual sounds.

Japanese is unlike other Altaic languages in that it possesses a very simple syllable structure; most syllables consist of just a vowel or a consonant followed by a vowel. Also, somewhat unusually among the Altaic languages, Japanese has only five vowels and a small number of consonants. As in other Altaic languages, the basic word order is subject–object–verb, and it uses suffixes to indicate the grammatical function of a noun, i.e. whether the noun is the subject (*-ga*) or an object (*-o, -ni*).

Perhaps the most careful attempt to link Japanese to Altaic proper and to Korean was that presented by Roy Andrew Miller (1971). Despite his efforts, few scholars readily accept the Japanese connection with Altaic, although the affiliation between Japanese and Korean has gained a wider consensus.

Korean is spoken by 65 million people worldwide, the vast majority of them living in North and South Korea. The earliest examples of written Korean used Chinese characters, a practice that continued into the

twentieth century. However, a Korean alphabet, called Hangul, was invented in the fifteenth century. Hangul is now the primary means by which Korean is written.

Chinese has had an obvious effect on Korean. In fact, more than half the Korean words have been borrowed from Chinese. Even so, the structure of Korean remains quite distinct from Chinese, and superficially at least, it shares many properties with Altaic proper. For example, the basic constituent order is subject–object–verb, it has no definite articles, and it exhibits 'vowel harmony'; i.e. vowels within a word tend to assimilate to one another.

A family relationship between Korean and Altaic proper was advanced by two leading Altaicists in the twentieth century: G.J. Ramstedt (1952–1957) and Nicolas Poppe (1965). The connection between Korean and Japanese has received more attention in the intervening years, sometimes being examined within the context of Altaic (for example, by Roy Andrew Miller [1972]) and sometimes independently of Altaic (for example, by Samuel Martin [1966, 1991]).

Mongolic languages are spoken in the central Asian regions of the People's Republic of China, the Russian Federation, and the Republic of Mongolia. Although there has been much debate regarding the internal structure of the family, it is perhaps most often depicted as having two primary branches. Western Mongolic consists of a single language called Moghul. The language, which is mutually unintelligible with all other Mongolic languages, is spoken by fewer than 200 people found in two villages in Afghanistan. Eastern Mongolic is broken down further into three sub-branches: Dagur (which consists solely of the Dagur language, spoken by around 60,000 people in China), Monghuor (a set of closely related languages spoken in China, all of which have been deeply influenced by Chinese and, in some cases, Tibetan languages), and Oirat-Khalka (the subbranch containing most Mongolic languages).

Within the Oirat-Khalka branch, the three most widely spoken languages are Oirat, spoken by approximately 300,000 people in the Kalmyk Autonomous Region in Russia; Buriat, spoken by 350,000 people in the Buriat-Mongol Autonomous Region in Russia; and Khalka, the official language of the Mongolian Republic, with more than 2.3 million speakers. It is this last language that most people refer to when they talk about Mongolian.

The oldest written records of Mongolian date back to the early thirteenth century and take the form of inscriptions, but the literary text known as *The Secret History of the Mongols* had appeared already in 1240. The script used for this early writing was borrowed from the Uighurs, a neighboring Turkic group that had been using the script for many centuries.

Like the other Altaic proper families, Mongolic languages typically use vowel harmony. The basic word order tends to be subject–object–verb, the grammatical function of nouns is signaled by suffixes, and there are no definite articles ("the").

Tungusic languages are spoken over a geographic expanse that goes all the way from the Arctic regions of the Russian Federation to China's Xinjiang Province. Most of these languages were traditionally spoken by small nomadic groups. As nomadism has come to an end in China and Russia, Tungusic speakers have Russified and Sinicized rather rapidly, and most Tungusic languages will probably start to die out in the near future.

Tungusic has traditionally been divided into two branches: Northern Tungusic and Southern Tungusic. The best-known member of the Southern branch is Manchu, which was the official language of China's last dynastic empire. Although the language was used for official purposes well into the twentieth century, very few fluent speakers remain. Sibe, another Southern Tungusic language, which, in fact, might be better labeled a dialect of Manchu, has 27,000 speakers in Xinjiang Province, making it the largest Tungusic language. The Northern branch of Tungusic consists of Even and a cluster of very closely related languages—Evenki, Negidal, Solon, and Orochen. Many scholars suggest that some of the Southern languages should be set off into a third branch of the family, Central Tungusic. The Central Tungusic languages, for example, Nanai, tend to be similar to Southern Tungusic in their pronunciation but similar to Northern Tungusic in the ways in which words are formed.

Only two Tungusic languages have written records that extend back before the twentieth century: Manchu and Jurchen. The Jurchen data date to the twelfth century, but its writing system mixes logographic symbols, which stand for concepts, with symbols for pronunciation, and it is still not completely deciphered. The Manchu script was derived from Mongolic script at the end of the sixteenth century. Today, we have a huge body of Manchu literature spanning the seventeenth to the twentieth centuries.

Tungusic languages are characteristically Altaic in having vowel harmony, case suffixes, and possessive suffixes. Also in line with other Altaic languages, they lack definite articles and have a weak distinction between nouns and adjectives.

The Turkic language family covers a large area from the Balkans and Asia Minor in the southeast to the Central Asian regions of both the Russian Federation and the People's Republic of China and up into the Arctic region of Russia. Turkic is split into two main branches: Bulgar and Common Turkic. The main language in the Bulgar branch is Chuvash, which is

spoken by about 1.5 million people in the Chuvash Autonomous Region of Russia. Chuvash, although grammatically similar to the rest of the Turkic languages, is marked by a series of historical changes in pronunciation that are hard to interpret in the context of Turkic and Altaic. Therefore, some scholars have suggested that Chuvash is actually a sister to proto-Turkic rather than a daughter. That is, instead of a branch of the Turkic family per se, Chuvash and the Turkic family may both be branches of a larger family unit.

Common Turkic is generally divided into four or five subbranches, most of which have languages spoken by more than a million people. The Eastern subbranch includes, among other languages, Uzbek (the national language of Uzbekistan, 15 million speakers) and Uighur (6.7 million speakers, mostly in China). The Southern subbranch includes Turkish (50 million speakers), Azeri (14 million speakers, half in Iran and half in Azerbaijan), and Turkmen (3 million speakers). The Northern subbranch includes Tuvan (200,000 speakers, mostly in Russian but also in Mongolia) and Yakut (300,000 speakers). The Western subbranch constitutes Bashkir (1 million speakers) and Tatar (5.5 million speakers). Many scholars separate out of Western Turkic a fifth subbranch called Central Turkic. Within Central Turkic, one finds the national languages of Kazakhstan, Kazakh (7.6 million speakers), and Kirghiz (2 million speakers).

The earliest written record of a Turkic language was found in a set of inscriptions in Mongolia dating to the early eighth century. These inscriptions were written in Orkhon Turkic and use what is called a runiform script. Some scholars have argued that the script is an invention of Orkhon Turkic speakers, although there is growing consensus that the script is actually of Semitic origin.

Turkic languages are characterized by vowel harmony. Typically, Turkic languages manifest a subject–object–verb word order and heavy use of suffixes. Nouns and adjectives are only weakly distinguished.

Speculation about some kind of genetic relationship among Mongolic, Tungusic, and Turkic stems back to the eighteenth century, and suggestions about the link between these languages and Korean and Japanese arose in the early nineteenth century. Despite a steady stream of scholarship with respect to the classification of these languages over the past 200 years, there is still widespread disagreement about the merits of an Altaic language family and, even granting its existence, which subfamilies should be contained therein.

With respect to Altaic, no one denies that there are a host of putative cognates, i.e. words that are quite obviously related. The issue is whether these cognates reflect shared ancestry or whether they have arisen through extensive borrowing among the languages.

Certainly, there is plenty of evidence that Altaic languages have been in contact for many centuries, and everyone involved in the Altaic debate agrees that much borrowing has taken place. The point of contention is whether these cognates should be dismissed en masse as borrowings or whether a subset of the cognates does in fact reflect a family association.

Adding fuel to the notion that putative Altaic cognates are not good indicators of a family relationship is the fact that among the basic vocabulary in the languages, for example, the lower numerals, body parts, and natural phenomena, the cognate evidence is not compelling. Typically, one expects better cognate evidence in these vocabulary realms.

One of the strongest pieces of evidence in favor of Altaic, or at least Altaic proper, is the first-person singular pronoun ('I'). In Tungusic languages, the nominative form of the pronoun is *bi,* but for all other cases, the stem is *min-*; e.g. Oroqen *min-ŋi,* 'my', is the genitive (possessive) first-person singular. Precisely the same contrast in stems is found widely in Mongolic. Hence, Buriat has *bi,* 'I', for the nominative first-person singular and *min-ni,* 'my', for the genitive. Although the correspondences in Turkic are not perfect, one does find similar patterns in some modern Turkic languages (e.g. Chuvash), and Old Turkic allowed variability in the pronominal stems; this, if not fully corresponding to the Tungusic and Mongolic patterns, shows enough similarity to suggest a connection among all three families.

References

Binnick, Robert I. 1987. On the classification of the Mongolian languages. Central Asiatic Journal 42. 178–95.

Johanson, Lars, and Éva Ágnes Csató, (eds.) 1998. The Turkic languages. London and New York: Routledge.

Kornfilt, Jaklin. 1987. Turkish and the Turkic languages. The world's major languages, ed. by Bernard Comrie. New York and Oxford: Oxford University Press.

Martin, Samuel E. 1966. Lexical evidence relating Korean to Japanese. Language 42. 185–251.

Martin, Samuel E. 1991. Morphological clues to the relationships of Japanese and Korean. Patterns of change, change of patterns: linguistic change and reconstruction methodology, ed. by Philip Baldi. New York: Mouton de Gruyter.

Miller, Roy Andrew. 1971. Japanese and other Altaic languages. Chicago: University of Chicago.

Patrie, James. 1982. The genetic relationship of the Ainu language. Honolulu: University of Hawaii Press.

Poppe, Nicholas. 1965. Introduction to Altaic linguistics. Wiesbaden: Otto Harrasowitz.

Ramstedt, G.J. [1952–1957.] Einführung in die altaische Sprachwissenschaft (Introduction to Altaic linguistics), 2 vols. Helsinki.

Street, John 1962. Review of Vergleichende Grammatik der altaischen Sprachen by Nicholas Poppe. Language 38. 92–8.

LINDSAY WHALEY

See also **Ainu; Chinese (Mandarin); Japanese; Korean; Tungusic**

Ambiguity

'Ambiguity' is the term for the existence of at least two separate, incompatible interpretations of a stretch of speech. Two kinds of ambiguity are recognized: if it is due to words, it is called 'lexical' ambiguity; if it is due to sentence structure, it is called 'structural' ambiguity.

Lexical ambiguity can be either 'polysemy' or 'homonymy'. Homonyms are distinct units in the vocabulary with identical pronunciation ('homophones') and/or spelling ('homographs'). Each of the homonyms in a set has its own meaning, for example, *bank* 'financial institution' vs. *bank* 'slope'. In this example, homophony and homography co-occur. But this need not always be the case; compare *peak* [pi:k] 'summit' and *peek* [pi:k] 'glance', which are homophones but not homographs. From a historical point of view, homonyms cannot be traced back to a common etymological origin, as with *bat:* the word for the flying mammal has a Swedish origin, the word for the implement is related to *battle*. From the contemporary viewpoint, homonyms have no shared meaning either. These two viewpoints are not watertight, however. A very competent speaker knows more about the meaning of a given pair of homonyms than a less competent speaker—but whose competence should be taken as the yardstick? Nor is it clear at which point etymological reconstruction should come to an end.

Multiple meanings distributed over several words with the same pronunciation is called 'homonymy'; multiple meanings within a single word is called 'polysemy'. The different meanings of a polysemous expression have a base meaning in common. Furthermore, the meanings of a polysemous term are often related by metaphor or metonymy, as in *point:* 'punctuation mark', 'sharp end', 'detail, argument', etc. Here, we observe several meanings associated with one word, the shared base meaning of which could be something like 'smallest unit'—either a concrete one ('punctuation mark', 'sharp end') or a metaphorical one ('detail, argument'). There are cases of idiosyncratic polysemy, such as *green* 'a certain color' and 'inexperienced', but there are also cases of systematic polysemy, as in the actual/dispositional distinction (as with *fast* in *this is a fast car*) or the distinction between a building and the institution housed in it (as with *school*).

Structural ambiguity has two subcases as well: 'attachment' ambiguity and 'scope' ambiguity. Attachment ambiguity refers to the possibility that the same sequence of words may be assigned different structures. In *The policeman observes the lady with the telescope*, the prepositional phrase *with the telescope* modifies either *the lady* (thus, the lady is a lady with a telescope) or *observes* (thus, the policeman observes by means of a telescope). The ambiguity of the sentence is an attachment ambiguity—the prepositional phrase may be 'attached' to or associated with different elements in the sentence.

Scope ambiguity refers to the possibility of assuming different logical forms of a sentence. An example is the sentence *Every man loves a woman,* which has two distinct readings: *for each man there is 'his' woman, and he loves her;* or alternatively *there is a specific woman who is loved by all the men.* With the first reading, *every man* 'has scope over' *a woman,* i.e. the sentence is primarily about 'every man'. With the second reading, it is the other way round, i.e. the sentence is primarily about 'a woman'.

Ambiguity, whether lexical or structural, must be distinguished from vagueness. A vague expression is imprecise, whereas an ambiguous expression has several precise meanings. An example of a vague expression is the predicate *red* in *Mary owns a red skirt.* A dark pink or a dark orange skirt would be borderline cases for this sentence, due to the intrinsic vagueness of *red*. Vagueness is 'intrinsic' in the sense that it has nothing to do with lack of knowledge (we know what the color *red* looks like). There is a close connection between vagueness and context dependence. The existence of borderline cases is not necessarily a bad thing, as different usages may be salient in different situations; cf. *Mary owns an expensive skirt*. Depending on the financial situations of Mary and the listeners, the vague predicate *expensive* will have quite different readings. In sum, vagueness may be defined as follows: a sentence is vague if—despite the knowledge of all the circumstances in a given situation—one cannot determine with certainty whether it is true or false. An expression is vague if it occurs in a sentence in such a way that it is responsible for the sentence's vagueness.

Vague sentences violate the principle of classical logic demanding that a sentence be either true or false. Thus, either there is a third truth value (or many) besides 'true' and 'false'—this is assumed in the many-valued logic approach (three-valued logics and fuzzy logics). Or there are gaps in truth-value assignment—this is the thesis of supervaluation semantics.

In sum, vagueness and ambiguity have different communicative statuses. Vagueness is part of the common understanding of the speaker and hearer, accepted

by both. With ambiguity, the situation is different. Usually, the speaker has a certain meaning in mind—there is no vague reading of an ambiguous sentence. Instead, any vague reading of an ambiguous sentence is a communicative breakdown, as ambiguous sentences are always in need of disambiguation.

References

Ballmer, Th.T., and M. Pinkal (eds.) 1983. Approaching vagueness. Amsterdam: North-Holland.

Lyons, J. 1977. Semantics. Cambridge: Cambridge University Press.

Pinkal, M. 1991. Vagheit und Ambiguität. Semantics, ed. by A.v. Stechow and Dieter Wunderlich. Berlin and New York: de Gruyter.

Rescher, N. 1969. Many-valued logic. New York: McGraw-Hill.

Williamson, T. 1993. Vagueness. London: Routledge.

Zwicky, A., and J. Sadock. 1975. Ambiguity tests and how to fail them. Syntax and semantics, Vol. 4, ed. by J. Kimball. New York: Academic Press.

MONIKA RATHERT

American Sign Language

American Sign Language (ASL) is a naturally occurring language with a syntax and culture distinct from spoken English. Geopolitically, ASL covers a large territory, serving between 100,000 and 500,000 signers in both Canada and the United States. ASL is not a 'universal sign language', nor does it have close etymological ties with British Sign Language, another distinct signed language. Instead, French Sign Language, or langue des signes française (LSF), has historically given ASL its closest linguistic ties.

Throughout early history deaf people lived isolated lives, possibly communicating with their hearing parents and siblings through the use of idiosyncratically created home signs. As world populations grew, larger metropolitan groups began to provide the contact for visual gestural dialects to emerge. In the late 1760s, the French priest Charles-Michel de l'Epée, concerned for the religious salvation of deaf Parisians, gathered a small group of deaf children together and began teaching them through the use of an invented manual French that he created as an educational counterpart to spoken French. This gathering in Paris is now recognized to be the first public school for the deaf in the world. The home signs and various Parisian dialects that the deaf children brought to their school merged with de l'Epée's 'methodological' signs and subsequently evolved into the language now known as LSF.

The reputation of the school and the emerging language flourished. In the early nineteenth century, a Protestant minister Thomas Hopkins Gallaudet was sent by wealthy American philanthropists to Europe to investigate a school for the deaf in London. Unable to obtain any assistance from England, Gallaudet visited France and eventually persuaded Laurent Clerc, one of the young deaf teachers from de l'Epée's school, to travel back to America with him. Together, they established the first American School for the Deaf in Hartford, Connecticut. The French language taught by Clerc mingled with the indigenous dialects brought in by his young American students from across the developing country. Mingling with a couple of fledging signed languages found in several small established deaf communities such as Martha's Vineyard in Massachusetts (descendants of British immigrants), and Henniker, New Hampshire, a new national signed language began to evolve.

In 1864, a liberal arts college now known as Gallaudet University was established in Washington, DC through the efforts of Abraham Lincoln and Edward Miner Gallaudet, Thomas Hopkins's son. Graduating deaf students, many of whom had previously attended Clerc's Connecticut school, returned to their native homes and frequently established or became responsible for establishing schools for deaf children in their respective states. The unique language mixture of Clerc's LSF signs and indigenous signs brought in or naturally created by Gallaudet students gradually evolved into modern ASL.

Spoken languages can usually trace their history back through many centuries. Even including its historical French ties, ASL has only 250 years of documented language use. Yet, ASL is already following a process of grammatical evolution that spoken languages follow. Cross-linguistic studies of the historical relationships between LSF and ASL reveal paths of lexical and grammatical change taking place in this young signed language.

One example can be found in the semantic changes that have occurred between the LSF nominal sign for 'coin' and a modern ASL discourse marker used to stop arguments or to change the subject (Wilcox 2000). The old LSF classifier noun for money, ARGENT (sign words are conventionally capitalized), semantically evolved into the sign for GIVE (donner), a verb in LSF that can also be found in modern ASL. The verb stem phonology of both 'give' signs has a common etymology, but diverge functionally. Following a path of lexicalization, one GIVE sign became an inflected verb of both literal and metaphorical nature; the other GIVE sign grammaticalized and extended into the word GIVE-IN-ARGUMENT. This verb is used metaphorically only and functions solely as a discourse marker.

Typical paths of grammaticalization in spoken languages reveal that language evolves from the mapping of literal objects onto nouns (nominalization) to semantic shifts of abstraction. The original concrete reference of a French money sign became increasingly abstract as it evolved into the metaphorical discourse marker in ASL. This example, plus the documented grammaticalization of modals such as WILL and MUST, demonstrate that ASL is following a linguistic path of subjectivity taken by many spoken languages.

ASL has an underlying s(ubject)–v(erb)–o(bject) sentence structure. In addition, it is strongly characterized by topic-comment structure. Topics act as discourse pivots to link previous information with the following and also indicates a shift in the point of view that the signer takes of the discourse events (Janzen et al. 1999). Topic marking in ASL is generally a marker with explicit discourse function. The following example also demonstrates the flexible word order of ASL:

[KIP BUY NEW COMPUTER]-topic, CURIOUS
'I wonder if Kip bought a new computer.'

Signed words consist of parts and can be broken down into components called parameters: the handshape, the location, the movement, and the orientation of the palm. These components exhibit both sequentiality and simultaneity when the signed words are put together, differing from spoken languages that are highly sequential. The iconic nature of ASL, with its strong link between form and meaning, is pervasive and is documented throughout its phonological and morphological analysis.

The morphology of ASL verbs is especially complex, more so than spoken English. The verb root is often expressed in the handshape and grammatical information is expressed in the movement, as Ronnie Wilbur describes:

> …in a multimorphemic verb sign, the movement can carry the information regarding the transfer of a theme-object (e.g. GIVE), the starting location can carry the source-agent-subject, the ending location can carry the goal-recipient-indirect-object, the handshape can carry the physical characteristics of the object itself (size and shape indicated by 'handle' classifier), and each piece of information is phonologically distinct from the others. (2000:216)

ASL has a large set of signs called classifiers that can be used to specify something without actually naming the specific referent. Classifiers can represent pronouns and describe in detail the size, shape, depth, or texture of an object, and they can function as a noun and indicate the location of that noun and its actions (Baker-Shenk and Cokely 1999). Classifiers are abundant in ASL and most signed languages.

Another critical component of ASL is the use of nonmanual signs. Grammar can be produced and distributed across the face, the head, and the shoulders. Adverbial and adjectival information can be produced on the mouth, tongue, and cheeks. For example, one visual gesture, 'th', is made with a slack jaw and a protruding tongue pushing through slightly separated lips, and indicates carelessness or incorrectness. Other nonmanual markers such as the eyebrows, the nose, head movements, and eyegaze can coarticulate with signed words or be produced in isolation. These nonmanuals mark wh-questions, yes/no questions, rhetorical questions, conditionals, negations, topics, and relative clauses. Rhetorical questions, uncommonly used in spoken English, are found quite frequently in ASL conversations. A rhetorical question is accompanied by a brow raise and a tilting of the head and often involves the use of the signs WHY, WHAT, WHO, HOW, and REASON. Nonmanual facial expressions can also indicate mood and affect.

Fingerspelling makes use of 26 handshape configurations that correspond to the English alphabet. Fingerspelling is more prevalent in ASL than in other signed languages of the world. It is primarily used to indicate proper nouns and technical terms. Fingerspelling is more than a sequence of canonical handshape-alphabet letter correspondence, however, since the articulatory movements of segments within the fingerspelled word influence each other. Forward (carry-over) and anticipatory assimilation affects the actual shaping of fingerspelled words, creating a fluid transition between letters that is prosodic and complex (Wilcox 1992).

Language borrowing involving fingerspelling takes place through the physical proximity of English and

ASL. Fingerspelled English words are often modified so that they fit better the structure of a sign (Battison 1978). For example, the word A-L-L can be fingerspelled letter by letter, or it can serve as a borrowed loan sign and will be restricted semantically in several ways. Instead of fingerspelling the three letters in the conical signing space in front of the body, a signer can hold up an imagined list with one hand while using the other hand to fingerspell the letters, literally running down the entire length of the arm to show 'all' of the items on the list. Or a two-handed loan sign can fingerspell an abbreviated A-L while sweeping the letters in front of the signer to designate everyone in the room including the signer. This particular lexicalized loan sign is highly productive and limits the semantic meaning according to specific modifications that are applied to the three letters.

The culture of deaf people endures not through artifacts such as food or clothing, but through the powerful use of the language of ASL. Historically, the hearing world has not condoned the language choice of deaf people. A defining event in 1880 still exerts influence on deaf people and their use of signed languages. The representatives of the World Conference for the Deaf in Milan, Italy, voted unanimously (except the American delegates) to ban signed languages from all the schools in Europe and replace it with an 'oral' method of instruction (Lane 1984). Even in America, this edict had a profound effect on the educational systems. Deaf teachers were fired or prevented from educating younger deaf children, teachers who signed were not promoted or respected, older students who might 'contaminate' the younger children were hidden in separate quarters, and curricula were changed to speech-based activities.

The most influential leader of the oral method of educating deaf children was Alexander Graham Bell, the inventor of the telephone. His great influence and wealth had a lasting impact upon the movement. Organizations that were founded by his proponents still work to oppress the use of ASL in the United States today. In the early 1900s, with schools no longer administered by deaf adults, the language was kept alive by the use of deaf children born of deaf people, a heritage unlike that found in spoken languages. Only 8–10% of deaf children are born to deaf parents who sign, which means that around 90% of the ASL-using deaf children learn it from their 'native' peers. The proponents of the oral method were relatively successful in keeping signed language out of the educational arena until the early 1960s when William C. Stokoe's (1965) seminal work on linguistic analysis of ASL appeared. His pioneering work ignited an explosion of research on ASL that now permeates linguistic inquiry. Because of the tenacity of ASL users, along with the backing of research that provides linguistic legitimacy, ASL is now finding rewarding acceptance in schools, universities, courtrooms, and social and business settings across America.

The past 200 years have witnessed a strong expansion of ASL throughout the world due to religious or educational agendas. Also, thousands of deaf students from around the world have attended Gallaudet University (the world's only liberal arts university for deaf people); thus, the signed language can be found throughout Asia, South America, Africa, and Europe — almost every country has deaf alumni who use ASL in addition to their own native signed language. Many deaf people have a working knowledge of ASL and the language is used freely at most of the international deaf conferences and workshops. However, its pervasive use of fingerspelling and marked initialization of English words has thus far prevented it from being accepted as an official signed lingua franca. Nevertheless, ASL users provide incentive to the world's remaining deaf populations to maintain their cultural and linguistic identities as their respective signed languages advance and evolve.

References

Armstrong, David F., Michael A. Karchmer, and John Vickrey Van Cleve. 2002. The study of signed languages: essays in honor of William C. Stokoe. Washington, DC: Gallaudet University Press.

Baker-Shenk, Charlotte, and Dennis Cokely. 1980. American sign language: a teacher's resource text on grammar and culture. Washington, DC: Clerc Books, Gallaudet University Press. (fourth printing, 1999).

Battison, Robbin. 1978. Lexical borrowing in American sign language. Silver Spring, MD: Linstok Press.

Janzen, Terry, B. Shaffer, and Sherman Wilcox. 1999. Signed language pragmatics. Handbook of pragmatics, ed. by Jef Verschueren, Jan-Ola Ostman, Jan Blommaert, and Chris Bulcaen, 1–20. Amsterdam: John Benjamins Publishing Company.

Klima, Edward S., and Ursula Bellugi. 1979. The signs of language. Cambridge, MA: Harvard University Press.

Lane, Harlan. 1984. When the mind hears. New York, NY: Random House.

Lane, Harlan, Robert Hoffmeister, and Ben Bahan. 1996. A journey into the deaf-world. San Diego, CA: DawnSignPress.

Padden, Carol, and Tom Humphries. 1988. Deaf in America: voices from a culture. Cambridge, MA: Harvard University Press.

Stokoe, William C., Dorothy Casterline, and Carl Croneberg. 1965. A dictionary of American sign language on linguistic principles. Washington, DC: Gallaudet College Press. Reprint, Silver Spring, MD: Linstok Press, 1976.

Wilbur, Ronnie. Layering of nonmanuals in ASL. 2000. The signs of language revisited: an anthology to honor Ursula Bellugi and Edward Klima, ed. by Karen Emmorey and Harlan Lane, 215–44. Mahwah, N J: Lawrence Erlbaum Associates.

Wilcox, Phyllis Perrin. 2000. Metaphor in American sign language. Washington, DC: Gallaudet University Press.

Wilcox, Sherman. 1992. The phonetics of fingerspelling. Philadelphia, PA: John Benjamins Publishing Company.

PHYLLIS PERRIN WILCOX

See also **British Sign Language; Grammaticalization**

Amharic and Ethiopian Semitic Languages

There are perhaps 11 Ethiopian and Eritrean Semitic (EES) languages now spoken, although the presence of dialect continua and lack of information on intelligibility between varieties makes a fully confident count difficult. Two other languages are known, but are now extinct. These 13 languages are (spelling of names of lesser known varieties varies in the literature):

1. Tigre, of western Eritrea, with dialects on the Red Sea coast and Dahlak islands;
2. Tigrinya, spoken in Eritrea and northern Ethiopia, but especially in Ethiopia's Tigray region;
3. Ge'ez, the ancient language known in epigraphy since about 500 BC, the language of the ancient kingdom of Axum centered on the modern Tigrayan city of the same name, unused as a mother tongue for many centuries, but continuing in regular use as the liturgical language of the Ethiopian Orthodox Christian Church;
4. Gafat, of the Gojjam region of central Ethiopia, extinct now for some 50 years;
5. Amharic, the world's most populous Semitic language after Arabic, with over 16 million speakers in Ethiopia, and long that nation's official language;
6. Argobba, near mutually intelligible neighbor to Amharic, in the northwestern Shewa region;
7. Soddo (also known as Kestane 'Christian'), spoken just west of Addis Ababa, with a dialect Dobbi (or Gogot);
8. Mesqan, south of Soddo, whose linguistic separateness from neighboring varieties is least confident;
9. Chaha and mutually intelligible varieties including Ezha, Gomara, Gura, and Muher, spoken south of Mesqan territory;
10. Inor and mutually intelligible varieties including Endegenya, Enner, Gyeta, Magar, and Mesmes (the latter probably now extinct), southern neighbors of the Chaha group;
11. Silte and mutually intelligible varieties including Enneqor, Ulbareg, and Walane, southeastern neighbors of the Inor group;
12. Zay, about Lake Zway some 70 km south of Addis Ababa; and
13. Harari (Adare), spoken in and about the ancient eastern Ethiopian city of Harar.

Some 76 languages are spoken in Ethiopia and Eritrea, in addition to the 11 Semitic languages, some 21 Cushitic languages, 23 Omotic languages, 20 Nilo-Saharan languages, and one as yet unclassified, Ongata. The presence of so many Semitic, Cushitic, and Omotic languages, three of the six subgroups of the Afroasiatic family (the other three Egyptian, Berber, and Chadic), suggests that the region is one of very early Afroasiatic settlement.

Amharic, with over 16 million mother tongue speakers and, according to the 1994 Ethiopian census, over five million second-language speakers, is by far the most important Ethiopian language, and the second most populous Semitic language after Arabic. Amharic is spoken throughout Ethiopia but particularly in urban areas, and is taught in all public schools. Regional varieties or dialects of Amharic are recognized in regions of Begemder, Gojjam, Wello, and Shewa. The Ethiopian capital city of Addis Ababa (in Shewa) is nowadays the focus of Ethiopian economic and cultural life, and Addis Ababa Amharic has naturally become the prestige dialect.

The numbers of speakers of most Ethiopian languages may be estimated from counts of the 1994 Census of Ethiopia. The Census, however, like much of the linguistic literature, grouped as 'Gurage' the six languages numbered 7–12 in the list, and provided separate figures for only the Soddo and Silte varieties. Amharic has three times more speakers than Tigrinya, with perhaps five and a half million including two million in Eritrea. Tigre may have a million speakers, all in Eritrea, followed by Ethiopian Silte (800,000), Soddo (200,000), and Chaha (perhaps 100,000). Speakers of the others are much fewer in number. The census reported Harari to have 22,000 speakers, and Argobba only 10,000.

The region is naturally characterized by bilingualism and multilingualism, often involving Amharic or Oromo, a Cushitic language (known in the old literature as 'Galla') perhaps as populous as Amharic, and more populous when Oromo speakers of Kenya are added. Research in the late 1960s showed, for example, that in Jimma town of western Ethiopia some 90% of speakers of languages other than Amharic knew Amharic, and 42% of Amharic speakers knew Oromo. Today, perhaps the majority of town and city-dwelling Ethiopians, except in largely Tigrinya-speaking Tigre province, are at least second-language speakers of

Amharic, and knowledge of the language certainly continues to spread throughout the region.

After a largely Tigrayan army overthrew the 1991 *derg* ('committee') communist government in Addis Ababa, whose predecessors had deposed the Amhara emperor Haile Sellassie in 1974, Ethiopian political power lay in the hands of Tigrinya speakers, who have encouraged considerable linguistic freedom of expression and identity including, rather controversially, the use of local-majority languages in public education. Oromo, backed by a nationalist movement, became a growing rival to Amharic-language dominance. With its many second-language speakers and long literary development, however, Amharic has largely continued its dominance in literary culture, education, and public life.

Written in Ge'ez, royal chronicles and religious writings are known from the fourteenth century, and shortly after this time writings in Amharic also began to appear, but especially in the seventeenth century. Publication in Amharic has increased steadily, and Ethiopian publications in Amharic today include writings of all kinds: newspapers, literary and news magazines, drama, novels, history, textbooks, and poetry. Amharic-language magazines are published in Europe and the United States to serve the Ethiopian expatriate populations there. Publication in Tigrinya as well as in Eritrea and Ethiopia's Tigray region has flourished.

Amharic and Tigrinya, and often nowadays other EES languages, are written in the Ge'ez or 'Ethiopic' writing system, an adaptation of the ancient South Arabian writing system, itself the southern adaptation of evolved Egyptian writing sometimes termed 'Sinaitic', whose northern adaptations include Aramaic, Arabic, and Hebrew writing (and Phoenician, perhaps that transformed by the Greeks into their alphabet). The epigraphic record suggests that during the fifth century reign of the Axumite king *Ezana,* Ethiopic writing was significantly modified to include representation of vowels, as regular extensions or shapings of its consonantal characters, approximately as early as in Indic *Kharosthi* writing.

Probably since ancient times, Amharic has spread in territory earlier populated by speakers of other, especially Cushitic, languages, and as a result Amharic has come to share a number of features with these other languages. Amharic has long borrowed from Ge'ez, a favored practice for the modern satisfaction of needs for philosophical, technical, and other new vocabulary. Borrowings from Italian entered the language in the Italian Colonial era, but today the principal source of new words is English, the language of Ethiopian and Eritrean secondary and higher education.

The obvious Semiticness of Amharic may be suggested by a comparison of a selection of five basic words of Hebrew, Classical Arabic, and Amharic.

	'all'	'die'	'eye'	'house'	'name'
Hebrew	*kōl*	*māt*	*ayin*	*bayit*	*šēm*
Arabic	*kull*	*māta*	*ayn*	*bayt*	*ism*
Amharic	*hullu*	*motə*	*ayn*	*bet*	*sɨm*

While a different selection of words might show no similarities, and many similarities that exist are far from obvious, the systematic prevalence of such comparisons as these confirms that Hebrew, Arabic, and Amharic are all descendant languages of that spoken by the first Semitic peoples.

Beyond lexical comparisons, Semitic languages, including EES, are particularly characterized by three-consonant roots filled out by vowels and affixes in word derivation (so-called 'root-and-pattern' morphology), and other characteristics including the presence of a set of subject-agreeing verb prefixes including *y-* for third-person masculine 'he', and *t-* for both second-person 'you' and third-person ('she') feminine:

	3m.sg.	3f.sg.	2m.sg.
Hebrew	*yi-kbad*	*ti-kbad*	*ti-kbad*
Arabic	*ya-kbud*	*ta-kbud*	*ta-kbud*
Amharic	*yɨ-kbəd*	*tɨ-kbəd*	*tɨ-kbəd*

The root (*kbd* heavy') and the prefixes are cognate; a difference is that the Amharic verbs are subjunctives, vs. the present tense of the Hebrew and Arabic.

Within the Semitic family of languages, Amharic and its kin in Eritrea and Ethiopia are typically grouped as 'South Semitic' with the ancient and modern Semitic languages of South Arabia, such as epigraphic Sabaen of Yemen and modern Soqotri of Soqotra island. The unity of EES is suggested by several features which they tend to share, including a non-finite verb conjugation for all but the last of verbs in sequence (Amh. *səbro* 'he having broken'), verbs expressed as compounds of particle + 'say' (*zɨmm alə* 'he was quiet'), and a special verb of presence (Amh. *alləhu* 'I am present').

EES languages are divided into two groups, North (Tigre, Tigrinya, and Ge'ez) and South (Amharic and the rest), which differ in a number of characteristic features, of which two are the presence of so-called 'broken' noun plurals in the North (for example, Tigrinya *sor* 'ox', pl. *'aswar*), formed by modification of root structure instead of affixation, and present-tense doubling of the second root consonant of triconsonantal verbs (compare Tigrinya *yɨsəbbɨr* 'he breaks' vs. Amh. *yɨsəbɨr*). The South EES languages are bifurcated as an eastern group of Amharic and Argobba plus the Silte-group, Zay, and Harari, against a western group of the Chaha-and Inor-group languages plus Mesqan and Soddo. The eastern group typically augments present-tense main verbs with a suffixed auxiliary verb, as in

Amharic main verb *yɨsəbr-al* 'he breaks' vs. subordinate *yɨsəbr.*

Argobba differs from Amharic in small but numerous ways, probably resulting in at least slight mutual nonintelligibility. The divergence of Amharic and Argobba may reasonably have resulted from the separation of these peoples upon the adoption of Islam by the Argobbas in about the sixteenth century. The typical divergence between pairs of EES languages is much greater, on the order suggested by 50% or so of cognates in a hundred words of core vocabulary between North and South EES languages, and about 60% between South EES languages.

Ge'ez, because it is the only anciently known EES language, has sometimes been considered the ancestor language from which all the others descend, as Latin is ancestor to the Romance or Italic languages, but this does not seem to be the case, because there are old linguistic features of the other languages unreasonably absent in Ge'ez, pervasive Ge'ez features unreasonably absent in the others, and because the diversity of the other languages seems too great to date only from the early Axumite times of likely Ge'ez dispersal.

Amharic, like other EES languages, constructs typical transitive sentences with the main constituent order subject–object–verb (SOV), in contrast to the VSO typicality of Semitic Arabic and Hebrew. An Amharic topicalized object can precede a subject, but the verb is rigidly final. Amharic, however, has a number of word-order characteristics inconsistent with the basic SOV type, including prepositions instead of postpositions (*bə-bet* 'in house', *bə-* 'in') and verb prefixal clause subordinators (*sɨ-tsəbr* 'when you break', *sɨ-* 'when'). Typical SOV characteristics of Amharic and EES languages are the linear precedence of adjectives, genitives, and relative clauses before their modified nouns. The syntactic peculiarities of EES are typically attributed to the influence of Cushitic-language neighbors, which are mainly of the consistent SOV type, with postpositions and case suffixing absent in EES. Amharic and most other EES languages usually case-mark only definite objects, Amharic with *-n*: *wɨša-w lɨǰ-u-n nəkkəsə* 'the dog bit the boy'.

Amharic and other EES languages have the phonological characteristic of a series of glottalized ejective consonants, as in Amharic *k'utt'a* 'anger' (with ejectives *k'* and *t'*). These differ from the so-called 'emphatic' velarized/pharyngealized cognate consonants of e.g. Arabic. Some EES languages preserve the Semitic pharyngeals *ħ* and *ʕ*. As in the other Semitic languages, the voiceless labial stops *p* and *p'* are marginal, or limited to loan words such as *ityopp'ya* 'Ethiopia', from Greek. Like most of the EES languages, Amharic has the seven-vowel system *i, e, a, ɨ, ə, u, o;* others have a five-vowel system with length.

References

Bender, M. L., J. D. Bowen, R. L. Cooper, and C. A. Ferguson (eds.) 1976. *Language in Ethiopia.* London: Oxford University Press.

Hetzron, Robert. 1972. Ethiopian Semitic: studies in classification (Journal of Semitic Studies monograph 2). Manchester: Manchester University Press.

Hetzron, Robert (ed.) 1997. The Semitic languages. London: Routledge.

Leslau, Wolf. 1978. English-Amharic context dictionary. Wiesbaden: Otto Harrassowitz.

Leslau, Wolf. 1987. Comparative dictionary of Ge'ez. Wiesbaden: Otto Harrassowitz.

Leslau, Wolf. 1995. Reference grammar of Amharic. Wiesbaden: Otto Harrassowitz.

GROVER HUDSON

See also **Afroasiatic; Arabic; Hebrew: Biblical; Hebrew: Modern; Semitic Languages; Writing Systems**

Analogical Change

In a general sense, analogy is the correspondence of two or more entities with respect to certain properties. In linguistics, the term 'analogy' is most often used to describe processes in which a new linguistic object is created by aligning an existing linguistic object A with another existing linguistic object B, which serves as a model. A and B usually have some properties in common. For instance, Early Modern English *pigs* and *kine* are both plural nouns but differ in the way the plural is formed: *-s* vs. *-n* + vowel change. The form *pigs*—respectively *horses, stones,* and the like—is considered the model here and the plural of *cow* is aligned with the plural of *pig* and thus, by analogy, *kine* becomes *cows.*

Sound change and analogical change have already been considered by Neogrammarians as the fundamental processes of language change. The two important

types of analogical change are analogical leveling and proportional analogy. Analogical leveling results in the reduction of allomorphy within paradigms. The paradigms of the Old English and Old High German verb meaning 'to choose' contain forms with a fricative [z] or [s], where other forms have an [r]. In both languages, the paradigms were leveled out by adopting the same consonant throughout the whole paradigm. While Modern English chose [z], Modern High German dropped the fricative in favor of [r].

Old Engl.		Mod. Engl.		Old High Germ.		Mod.H.G.
cēo[z]*an*	>	*choo*[z]*e*	PRES	*kiu*[s]*an*	>	*kü*[r]*en*
cēa[s]	>	*cho*[z]*e*	PAST-(SG)	*ko*[s]	>	*ko*[r]
cu[r]*on*	>	*cho*[z]*e*	PAST-(PL)	*ku*[r]*un*	>	*ko*[r]*en*
(*ge-*)*co*[r]*en*	>	*cho*[z]*en*	PAST-PART	(*gi-*)*ko*[r]*an*	>	*geko*[r]*en*

The above example also illustrates the interaction of sound change and analogical change. The alternation of [s] and [z] results from a sound change referred to as 'Verner's Law' and [r] is the result of a subsequent change from [z] to [r] in certain positions ('rhotacism'). Sound change is conditioned by phonetic factors and is not sensitive to the needs of grammatical and semantic functioning. It often obliviates the principle 'one form—one meaning' ('Humboldt's universal'). Analogical leveling is always conditioned by nonphonetic factors and often 'fixes the damage' caused by sound change.

In contrast to sound change, analogy has been considered to be an unsystematic process. This has given rise to the formulation of 'Sturtevant's Paradox': sound change is regular but creates irregularity, whereas analogy is irregular but creates regularity. However, in some cases analogy is fairly systematic. For instance, [z]–[r] alternations of the above type have been leveled out in English quite systematically, as within the paradigms of *loose, freeze, and rise*. There are only very few cases where the [r] has been retained (e.g. *was* vs. *were*).

By proportional analogy—also called 'analogical extension' or 'four-part analogy'—an already existing morpheme or relation becomes generalized to other linguistic forms. In most cases, proportional analogy creates or assimilates derived forms on the basis of another morphological pattern. Proportional analogy proceeds like the solution of an equation of the form:

A : B = C : X solve for X
stone : *stones* = *cow* : X   X = *cows* (< *kine*)

The solution of this equation with respect to the phenomenon mentioned above yields the replacement of *kine* by *cows* as the plural form of *cow*. As with analogical leveling, proportional analogy is not fully systematic: the analogy did not carry over to words like *foot*: *feet* / **foots*.

Other less productive forms of proportional analogy are backformation and hypercorrection. With backformation, it is the morphologically less complex item that is created. In Early Modern English, *pease* was a mass noun, which ended in a sequence that resembled a plural morpheme. Its reanalysis led to the creation of a singular form *pea* by dropping the final fricative. Backformation also occurs as a word-formation process.

bean   : *beans* [-z] = X   : *pease* [-z]
                                                   (X = *pea*)
depression : *depress*   = *aggression* : X
                                                   (X = *aggress*)

Hypercorrection consists in imposing a pattern that relates a colloquial low-prestige variant and a standard higher-prestige variant onto another word presumed to be low prestige. Thereby, an allegedly high-prestige variant of the word is created:

fella' : *fellow*   = *umbrella* : X   (X = *umbrellow*)

Other phenomena that have been discussed under the heading of 'analogy' are blending, contamination, reanalysis, and folk etymology. Blending or 'portmanteau formation' consists in combining parts of existing words where these parts are usually not morphemes. Examples are *brunch* from *breakfast* and *lunch* or *infotainment* from *information* and *entertainment*.

In the case of contamination, a similar fusion of two words occurs but the resulting expression retains the meaning of one of the source words. For example, the [t] of Proto-Indo-European *pətēr* should have turned up as [d] by Verner's Law and a subsequent sound change, yielding *fader*. Instead, contaminated by the regularly derived *brother,* we get *father*.

In reanalysis, a deviant structure is assigned to an expression. The words *napkin* and *apron* both derive from French *nape* 'cloth'. In the second case, the Middle English sequence that results from the combination of the word with an indefinite article *anapron* was reanalyzed as *an apron*, interpreting *n* as part of the article. A related phenomenon is folk etymology. Sometimes, words change on the basis of what a speaker assumes is the etymological origin of the word. For instance, French *crévisse* should occur in English as something like *crevisse*, but appears as

crayfish because the second part got misinterpreted as referring to some kind of fish or aquatic creature.

Theoretical approaches to analogical change manifest themselves in attempts to construct a theory of analogical processes, which is explicit and restrictive and unifies the different forms of analogy (e.g. within generative phonology; cf. Kiparsky 1982). Some researchers formulated a number of principles (sometimes called 'laws') in order to explain why and how analogical changes occur (e.g. Kuryłowicz 1949, Mańczak 1980). These principles are usually subject to counterexamples and should therefore rather be understood as tendencies (cf. McMahon 1994). Among these principles were the following: (i) Analogy proceeds from a basic to a derived form or from a shorter to a longer form. Counterexample: backformations. (ii) More overt and complex markers are favored by analogy. For example, the Old High German plural forms of *Gast* 'guest' and *Topf* 'pan / pot' are *Gästi* and *Topfa*. By analogy *Topfa* adapted to the form *Gästi,* which has the more complex plural marking (umlaut and suffix) yielding *Gäste* and *Töpfe* in Modern High German. Counterexample: Old English *lang* : *leng-ra* turned into *long* : *long-er* analogous to *warm* : *warm-er*, not into *long* : **leng-er*. (iii) In processes of proportional analogy, it is usually the more productive pattern that serves as a model. Counterexample: extension of strong verb forms to weak ones as in the change from *dived* to *dove* as past tense of *dive* analogous to *ride* : *rode*. (iv) Allomorphy gets reduced by analogy (in particular by analogical leveling). Counterexample: the change of the stem *Topf-* to *Töpf-* in German mentioned above. (v) When a new analogical formation becomes accepted (e.g. *brother*) and the older form also survives (*brethren*), the analogical form takes over the basic meaning. Counterexample: blendings.

Analogy has also been considered an important tool in other branches of linguistics like theories on language acquisition, syntax, and orthography: (i) Analogy is assumed to be at work when children overregularize morphological forms, like *goed* instead of *went*. (ii) An example of syntactic analogy is the change of impersonal constructions like *me hungreth*, *me thinks* to the more widespread personal constructions, i.e. to *I hunger*, *I think*. (iii) Modern High German *lieb* 'kind / nice' orthographically reflects a Middle High German diphthong, which has been monophthongized to [i:] in Modern High German. Words like *Wiese* 'lawn' adopted this spelling by proportional analogy, although the Middle High German spelling and pronounciation *wise* showed a monophthong.

References

Antilla, Raimo, and Warren A. Brewer. 1977. Analogy: A basic bibliography. Amsterdam: Benjamins.

Hock, Hans Heinrich. 1991. Principles of historical linguistics, 2nd revised and updated edition. Berlin, New York: de Gruyter.

Hock, Hans Heinrich, and Brian D. Joseph. 1996. Language history, language change, and language relationship. An introduction to historical and comparative linguistics. Berlin, New York: Mouton de Gruyter.

Kiparsky, Paul. 1982. Explanations in phonology. Holland: Dordrecht / Foris: Cinnaminson.

Kuryłowicz, J. 1949. La nature des procès dits >analogiques<. Acta Linguistica Hajniensia 5. 15–37 [Reprint in: Hamp, Eric P., F.W. Householder and Robert Austerlitz, editors, Readings in linguistics, 2nd edition. Chicago: Chicago University Press, 1966].

Mańczak , W. 1980. Laws of analogy. Historical morphology, ed. by Jacek Fisiak. The Hague, Paris, New York: Mouton.

Marle, J. van. 1985. On the paradigmatic dimension of morphological creativity. Foris: Dordrecht.

McMahon, April M. S. 1994. Understanding language change. Cambridge: Cambridge University Press.

STEFAN ENGELBERG

See also **Comparative Method; Historical Linguistics; Inflection and Derivation; Language Change; Morphology**

Anaphora

Linguistic expressions may have their semantic interpretation determined solely by their lexical content (and the way these lexical items are syntactically arranged), or they may have all or part of their interpretation fixed by a relation to another linguistic expression. Examples of the first type in the following examples include phrases like *the tall woman, bought a car,* and *no tall man.* Examples of the second type include *herself (*see (1)*)*, *she/her* (on the construal whereby *she/her* refers to the same person as *the tall woman*—see (2) and (3)), *his* (on the construal whereby its meaning is determined by

no tall man—see (4)), and the unpronounced verb phrase in (5).

(1) The tall woman hurt herself.
(2) The tall woman realized that she was tired.
(3) The tall woman ran the risk of bumping her head on the low lintel.
(4) No short man runs the risk of bumping his head on a high lintel.
(5) Mary bought a car, and Stanley did ... too.

The study of the syntactic configurations in which the antecedent–dependent pair may occur, and of the types of semantic relation they may bear to each other, is called *anaphora*.

The most widely studied cases of anaphora in generative grammar are those involving nominal expressions like the reflexive in (1) and the pronouns in (2)–(4).

Semantically, anaphoric dependencies between nominals divide into at least two types. When the antecedent refers to an entity in the world (as in (1)–(3)), the relation is that of *coreference*—the dependent refers to whatever the antecedent does, and so they *corefer*. When the antecedent is quantificational, as in (4), the relation is that of *bound-variable anaphora*. *No short man* does not, strictly speaking, refer directly to an entity. Rather, it quantifies over the things that are men, and *his* acts as a bound variable. The sentence essentially means 'no thing x which is a man is such that x runs the risk of bumping x's head… .'

Nominals divide into three types with respect to the syntax of anaphora. *Anaphors* are nominals that must have an antecedent; these include both reflexives (see (1)) and reciprocals, e.g. *each other, one another* in English. *Pronominals*, like *she, her,* and *his,* may have antecedents (as on the most salient interpretations of (2)–(4)), but they need not. Each of (2)–(4) may be interpreted with the pronoun referring to some unmentioned person. Such interpretations typically involve either some kind of ostension (e.g. pointing at the intended referent), or prominence of the referent (e.g. the speaker at a conference). The third type of nominal is those that do not have antecedents at all, like *the tall woman, a low lintel, a car,* and *no tall man*. These are termed *R-expressions* (from 'referring expression').

Each of these three types of nominal has a distinct syntactic distribution with respect to other nominals with which it has an anaphoric relation. These syntactic conditions form the center of Noam Chomsky's Binding Theory (1980s). Each is critically dependent on a structural relationship between two elements called *c-command*. Without going into technical details, one can roughly say that one element c-commands another if it is (a) in a structurally more prominent position and (b) not contained within the same structural unit as the other.

Chomsky's Binding Theory is based on three constraints:

Condition A: an anaphor must be c-commanded by an antecedent within its binding domain.

Condition B: a pronoun must *not* be c-commanded by an antecedent within its binding domain.

Condition C: an R-expression must *not* be c-commanded by any nominal of which it is the antecedent.

In (the majority dialect of) English, the binding domain for an anaphor is the smallest argument phrase containing it and another c-commanding N(oun) P(hrase)—usually the clause. The binding domain for a pronominal is the smallest argument phrase containing it. Grammaticality judgments below (with ungrammaticality indicated by an asterisk) are given for the coreference reading of the antecedent–dependent pair; thus (10) is ungrammatical if *her* and *Mary* corefer).

(9) Mary hurt herself.
(10) * Mary hurt her.
(11) * Mary said that Dan admires herself.
(12) Mary said that Dan admired her.

In (9), *herself* is a reflexive, i.e. a syntactic anaphor. Its antecedent *Mary* is the c-commanding subject of the clause; hence, Condition A is satisfied. In (10), the pronoun *her* is contained in the same clause as the c-commanding antecedent *Mary*, which violates Condition B and makes the sentence ungrammatical. In (11), the reflexive *herself* does not have an antecedent within the binding domain (the embedded clause introduced by *that*), violating Condition A. (12) is grammatical, because the antecedent *Mary* of the pronoun *her* is not located within the same clause. Thus, Conditions A and B are basically *locality* constraints on anaphors and pronominals: anaphors must be close to their antecedents (the antecedent must be local), and pronominals cannot be (the antecedent must not be local).

No locality effects are obtained for Condition C. Coreference is blocked both in the local case (15) and in the long-distance case (16), where the pronoun and R-expression are separated by several clause boundaries:

(15) * She admires Mary.
(16) *She thinks that I said that Dan admires Mary.

In both cases, the main clause subject *she* c-commands *Mary*. If *Mary* is interpreted as the antecedent for *she*, Condition C is violated—the sentence is ungrammatical.

Since c-command plays a critical role, it appears that Chomsky's binding constraints, although they govern a semantic relation, are crucially syntactic in nature. Thus, the binding conditions are a prototypical example of the tight interplay between syntax and semantics. Examination of the parallel conditions in

other languages reveals that the definition of binding domain varies substantially across languages. The study of such a variation has led to detailed examination of dozens of languages, and has also been the subject of a number of studies in the first- and second-language acquisition literature.

Returning to bound-variable anaphora, this has a more restricted syntactic occurrence than coreference.

(17) John/every man said that he was elegant.
(18) [The woman who danced with John/*every man] said that he was elegant.

Coreference may occur where the antecedent c-commands the pronoun, as in (17), where *Every man* is itself the subject of the sentence. When there is no c-command relation as in (18), where *every man* is deeply embedded in the subject phrase, the pronoun cannot have an anaphoric relation to *every man*. An anaphoric interpretation of the pronoun would thus be ungrammatical. Notice that both sentences are grammatical when *every man* is replaced by *John,* because *he* and *John* would simply corefer, i.e. *he* would not be a bound variable.

References

Barss, Andrew (ed.) 2002. Anaphora: a reference guide. Oxford: Blackwell.

Fiengo, Robert, and Robert May. 1994. Indices and identity. Cambridge, MA: MIT Press.

Higginbotham, James. 1983. Logical form, binding, and nominals. Linguistic Inquiry 14(3). 395–420.

May, Robert. 1984. Logical form. Cambridge, MA: MIT Press.

Reinhart, Tanya, and Eric Reuland. 1993. Reflexivity. Linguistic Inquiry 24. 657–720.

ANDREW BARSS

See also **Chomsky, Noam**

Anatomy of the Articulatory System

Speech sounds are produced by moving air in and out of the body. To accomplish this, a number of organs in the chest, neck, and head are used. An understanding of these anatomical organs and their functions is a prerequisite for phonetics, the study of speech sounds.

The air pathway (Figure 1) consists of the lungs, larynx, pharynx (throat), oral cavity (mouth), and nasal cavity. The lungs start the process of speech production by pushing air upward. The vocal folds, which are located in the larynx behind the Adam's apple, may vibrate, causing the air that flows through them to vibrate as well. The airstream is then modified by the vocal tract (Figure 4)—the pharynx and the oral and nasal cavities. By altering the shape of the vocal tract, a large variety of sounds can be produced.

The lungs sit inside the rib cage. Expanding the rib cage causes the lungs to expand, and air is drawn in; when the rib cage and lungs are compressed, air is expelled. The lungs are elastic, like a sponge; if their size is changed, this elastic property tends to return the lungs to their resting size. The two lungs are independent. If one is injured, the other one still functions.

Tiny tubes within the lungs, called *alveoli,* combine repeatedly to form larger tubes, ultimately resulting in a single tube for each lung, called a *bronchus.* The two bronchi in turn join to form the *trachea* (windpipe). The top of the trachea is just behind the notch at the top of the breastbone. Lying behind the trachea, the *esophagus* is a tube that carries food to the stomach.

Various muscles of the chest expand and contract the ribcage. The *external intercostal* muscles lift the ribcage, inflating the lungs, and the *internal intercostals* pull the ribcage down, deflating the lungs. The *abdominal muscles* also help to deflate the lungs. The *diaphragm* is a large dome-shaped muscle lying just below the lungs; by contracting, the diaphragm moves downward, increasing the capacity of the lungs. The elasticity of the lungs resists expansion and contraction, attempting to return the lungs to their resting size.

The *larynx* (Figures 2 and 3) is a complex structure, cylindrical in shape, composed of several cartilages. It is found above the trachea. Within the larynx, the ring-shaped *cricoid cartilage* sits immediately atop the trachea. Above the cricoid cartilage is the plough-shaped *thyroid cartilage.* The forward external point of the thyroid cartilage is easily identified as the *Adam's apple.* The back of the thyroid cartilage has two horns pointing down and two pointing up. The two lower horns attach to the cricoid cartilage. The thyroid cartilage provides a shield for the vocal folds, which are attached behind the Adam's apple. The two *arytenoid cartilages* are small pyramid-shaped cartilages sitting
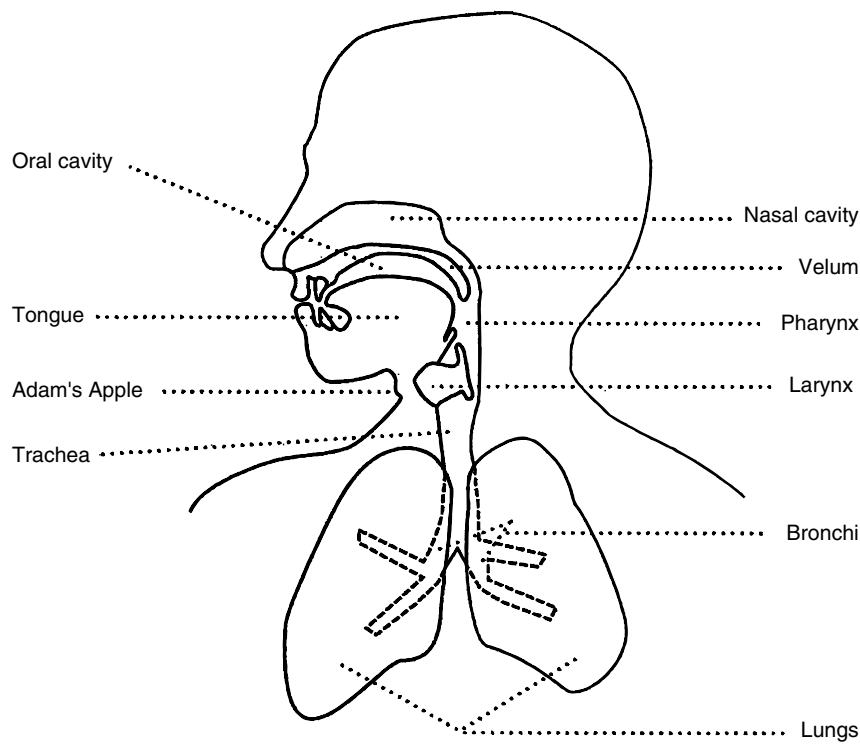
Figure 1. Vocal Organs.
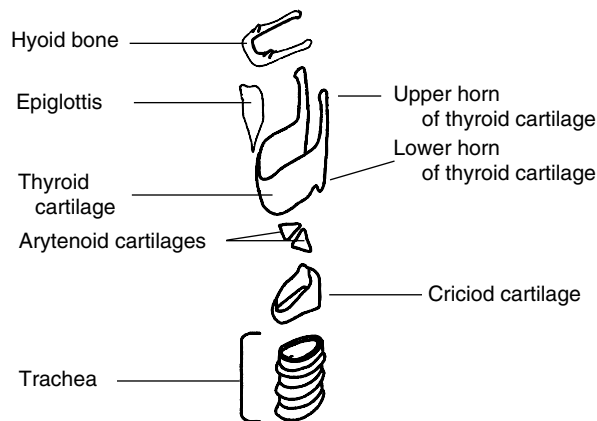
Figure 2. Larynx.
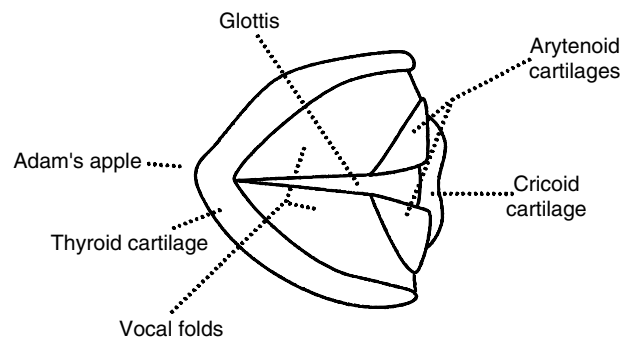
Figure 3. Larynx Overhead.

on top of the cricoid cartilage at the back on either side (see Figure 3).

Crucial to speech are the *vocal folds.* These are two horizontal shelves of muscle and ligament. The front part of both vocal folds is attached to the thyroid cartilage. Each vocal fold is attached at the back to one of the two arytenoid cartilages. The vocal folds have a triangular space between them called the *glottis,* through which the air passes upward from the trachea. The vocal folds assume a variety of positions to affect the airstream coming from the lungs. The different positions create the glottal stop as well as different vocal qualities—voiced, voiceless, creaky, breathy, falsetto. In the past, the vocal folds were often called vocal cords; they are, however, solid structures and not cord shaped. The attachment of the horns of the thyroid cartilage to the cricoid cartilage allows the thyroid cartilage to rock backward and forward, changing the tension of the vocal folds; this is the basic mechanism for varying the pitch of the voice (see Figure 3).

The *epiglottis* (Figure 2) is a spoon-shaped cartilage in the pharynx that forms something of a hood over the main part of the larynx. The horseshoe-shaped *hyoid bone* (Figure 2) resides behind the chin; it supports the muscles of the tongue above it, and it also stabilizes the muscles of the larynx from above (see Figure 2).

The *pharynx* is the name for the throat. It is a vertical tube connecting the larynx with the oral and nasal cavities. The rear wall of the pharynx is the *pharyngeal*

51
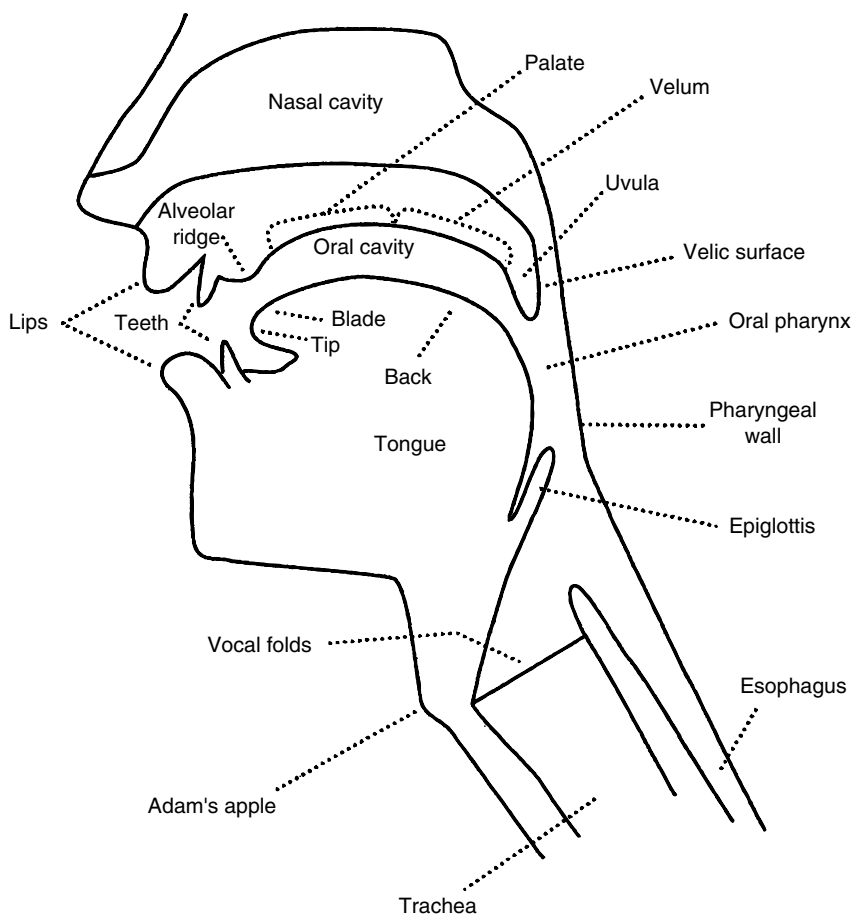
Figure 4. Vocal Tract.

wall. The root of the tongue forms the forward wall of the pharynx.

The *oral cavity*, or mouth, is extremely important in the production of speech sounds. The rear of the oral cavity connects to the pharynx. The *upper articulators* are the upper lip, the upper teeth, the upper surface of the mouth, and the back pharyngeal wall (see Figure 4). The *lower articulators* are the lower lip, lower teeth, and tongue. Sounds are formed by moving a lower articulator toward an upper articulator to form a complete or partial obstruction to the airflow.

The lips and teeth are familiar. A bumpy area known as the *alveolar ridge* lies just behind the upper teeth. The hard palate is a smooth bony structure behind the alveolar ridge. In phonetics, the hard palate is normally referred to simply as the *palate*. Behind the palate lies the *velum,* or soft palate. The back of the velum narrows at the back to a thin structure called the *uvula*. The *pharyngeal wall* is considered to be an upper articulator (see Figure 4).

The *lower articulators* are the lower lip, the lower teeth, and the tongue. The *tongue* is the most important of the lower articulators; it is a large, complex organ. Although the surface of the tongue is continu-
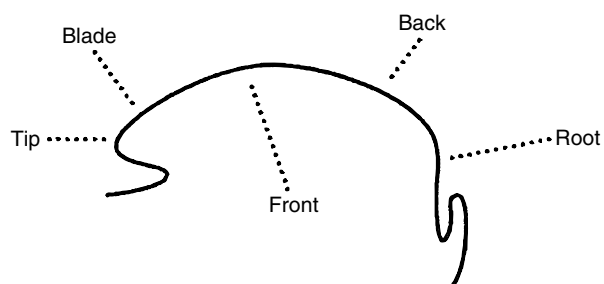
Figure 5. Tongue.

ous, without clear landmarks, phoneticians divide it into five parts (see Figure 5).

The *tip*, or *apex*, of the tongue is its foremost part. Just behind the tip of the tongue is a small surface called the *blade*, or *lamina*. The middle portion of the tongue is known as the *front*. It is not literally the front of the tongue, but the surface behind the blade. The rear portion of the horizontal surface of the tongue is called the *back* or *dorsum*. The *root* of the tongue is the rear vertical surface facing the pharyngeal wall (see Figure 5).

The top of the pharynx opens into the *nasal cavity.* The upper surface of the velum is known as the *velic*

surface. The velum can move up or down to close off the passageway to the nasal cavity.

References

Catford, J.C. 1977. Fundamental problems in phonetics. Bloomington, IN: Indiana University Press.

Catford, J.C. 1988. A practical introduction to phonetics. Oxford: Oxford University Press.

Gimson, A.C. 1980. An introduction to the pronunciation of English. London: Edward Arnold.

Hardcastle, William J., and John Laver (eds.) 1997. The handbook of phonetic sciences. Oxford: Blackwell.

International Phonetic Association. 1999. Handbook of the International Phonetic Association. Cambridge: Cambridge University Press.

Kahane, Joel C., and John W. Folkins. 1984. Atlas of speech and hearing anatomy. Columbus, OH: Charles E. Merrill.

Ladefoged, Peter. 1993. A course in phonetics. Ft. Worth, TX: Harcourt Brace Jovanovich.

Ladefoged, Peter. 2000. Vowels and consonants: an introduction to the sounds of languages. Oxford: Blackwell.

Ladefoged, Peter, and Ian Maddieson. 1996. Sounds of the world's languages. Oxford: Blackwell.

Laver, John. 1994. Principles of phonetics. Cambridge: Cambridge University Press.

Minifie, Fred, Thomas J. Hixon, and Frederick Hixon. 1973. Normal aspects of speech, hearing, and language. Englewood Cliffs, NJ: Prentice-Hall.

Pullum, G.K., and W.A. Ladusaw. 1996. Phonetic symbol guide. Chicago: University of Chicago Press.

Rogers, Henry. 2000. The sounds of language: an introduction to phonetics. Harlow, England: Pearson.

Zemlin, W.R. 1988. Speech and hearing science. Englewood Cliffs, NJ: Prentice-Hall.

HENRY ROGERS

See also **Speech Production**

Anatomy of the Auditory System

The auditory system consists of three main sections: the outer ear, the middle ear, and the inner ear. The structure and function of each of these systems will be discussed briefly in the next sections.

Outer Ear

The purpose of the outer ear is to gather sounds from the environment and to funnel them into the auditory system. The outer ear begins at the auricle (also *pinna*) and ends at the eardrum (also *tympanic* membrane). The auricle is mainly composed of elastic cartilage (except for the earlobe) and aids front-to-back localization. It leads to the opening of the ear canal (or external auditory canal). This canal is, approximately, 9 mm high, 6.5 mm wide, and 2.5 cm long, and provides some protection of the eardrum against foreign bodies. The eardrum is a smooth, nearly transparent, membrane of about 80 mm^2 in area. Although it is extremely thin (about 0.07 mm), it is constructed of three layers. It is attached to the handle of the hammer (malleus), the first of three tiny bones (ossicles) in the middle ear.

Middle Ear

The middle ear is an air-filled cavity that is connected to the nose cavity via the eustachian tube. The eustachian tube provides for the aeration and drainage of the middle ear system and makes it possible for air pressure to be the same on both sides of the eardrum (also while yawning or swallowing). While we are awake, the eustachian tube opens approximately once per minute; during sleep it opens, on average, once every 5 min. The eardrum is coupled to the oval window, an entrance to the inner ear, through a chain of three tiny bones, the hammer (malleus), anvil (incus), and stirrup (stapes). Sound waves traveling in the external auditory canal cause the eardrum to vibrate. This vibration is passed on to the middle ear, where a pressure increase is needed because of the mismatch in density and compressibility between air at the eardrum and fluid at the oval window. As the effective surface area of the eardrum is 55 mm^2 and that of the stapes footplate connecting to the oval window is about 3.2 mm^2, the vibratory pressure at the eardrum is funneled and increased by a factor of 17 (55/3.2) at the oval window. Pressure at the oval window is furthermore increased by the three bones acting as a lever system and by the conical shape of the eardrum.

Inner Ear

Although the inner ear is as tiny as a bean, it contains thousands of moving parts and is responsible for sending information to the brain regarding hearing and

balance. The primary auditory organ of the inner ear is the cochlea (latin for *snail*). It is about 1 cm wide, and 5 mm from base to apex in man, makes about two and a half coils, and is filled with fluid. The cochlear coil is divided into three channels: the scala vestibuli, scala media, and scala tympani. The scala vestibuli and scala tympani are filled with perilymph and communicate via an opening at the apex known as the helicotrema. The central duct, the scala media is filled with endolymph and bounded by two membranes, Reissner's membrane (top) and the basilar membrane (bottom), that form a wedge-shaped partition. On the basilar membrane (about 35 mm long), along the full length of the scala media, lies the organ of Corti, which contains the actual receptor cells that are responsible for hearing. On the inner side of the tunnel of Corti, a single row of inner hair cells, approximately 3,400 in total, are lined up side by side along the entire length of the organ of Corti (35 mm). On the outer side of the tunnel of Corti are three rows of outer hair cells, approximately 13,400 in total in humans. Each hair cell has 'hairlike' projections, called stereocilia, projecting from their apical surface. About 90–95% of the fiber of the auditory nerve connect to the inner hair cells. Many outer hair cells converge to a single auditory nerve fiber, while each inner hair cell may excite up to 20 auditory nerve fibers. The outer hair cells act to enhance the selectivity by changing the mechanical properties of the basilar membrane. If the outer hair cells are selectively damaged, there is a loss of sensitivity. The hair cells and nerve fibers are held in place by supporting cells.

The motion of the footplate of the stapes in the oval window moves the fluid of the inner ear. This air vibration, magnified by the middle ear, creates a pressure wave in the fluid, which distorts the basilar membrane in the scala media. Since the fluids of the cochlea are incompressible, the displaced fluid, in turn, causes an outward displacement of the round window. This produces a progressive traveling wave on the membrane from base to apex. Traveling waves produced by high-frequency sounds (shorter wavelength) show maximum displacement near the base of the cochlea, while those produced by a low-frequency sound (longer wavelength) come to a peak near the apex. Arching over the hair cells is the tectorial membrane. The differential motion of the basilar and tectorial membranes results in a shearing motion of the stereocilia of the hair cells. This bending of the hair cells produces an electrical discharge in the cochlear portion of the VIIIth nerve. A chemical is released at the base of the hair cell when the stereocilia are sheared. Although the mechanics of the organ of Corti are very complex as a result of different motions of the basilar membrane (up and down, side to side, etc.), the aforementioned briefly describes how mechanical motion is converted into neural activity in the organ of Corti. The size of the electrical response of the cochlea is directly related to the extent to which the hair cells are sheared.

References

Durrant, J.D., and J.H. Lovrinic. 1984. Bases of hearing science, 2nd edition, 276pp. Baltimore, MD: Williams & Wilkins.

Lim, D.J. 1980. Cochlear anatomy related to cochlear micromechanics. A review. Journal of the Acoustical Society of America 67. 1686–95.

Lim, D.J. 1986. Functional structure of the organ of Corti: a review. Hearing Research 22. 117–46.

Lipscomb, D.M. 1976. Mechanisms of the middle ear. Hearing disorders, ed. by J.L. Northern, 78–88. Boston: Little, Brown.

Maue-Dickson, W. 1981. The inner ear—prenatal development. Medical audiology, ed. by F.N. Martin, 220–35. Englewood Cliffs, NJ: Prentice-Hall.

Pickles, J.O. 1988. An introduction to the physiology of hearing, 2nd edition, 367pp. New York: Academic Press.

Yost, W.A., and D.W. Nielsen. 1985. Fundamentals of hearing: an introduction, 2nd edition, 269pp. Orlando, FL: Holt, Rinehart & Winston, Inc.

ASTRID VAN WIERINGEN

Ancient Egyptian

Ancient Egyptian is a branch of the Afroasiatic family (also called Hamito-Semitic in traditional comparative linguistics) attested in Egypt from 3000 BCE to 1300 CE. Within its family, it shows the closest relationships to Cushitic, Semitic and Berber. There are two main stages in the development of the Egyptian language (cf. Loprieno 1995):

(1) Earlier Egyptian (3000–1300 BCE), further subdivided into:

(a) Old Egyptian (3000–2000 BCE: Old Kingdom/First Intermediate Period);

(b) Middle or Classical Egyptian (2000–1300 BCE: Middle Kingdom/XVIII Dynasty; used in religious texts until the Greco-Roman period);

(2) Later Egyptian (1300 BCE–1300 CE), further subdivided into:

(a) Late Egyptian (1300–700 BCE: Dynasty XIX/Third Intermediate Period);

(b) Demotic (seventh century BCE to fifth century CE: originally the language of administration during the Late Period, used also in literary texts);

(c) Coptic (fourth to fourteenth century CE: the language of Christian Egypt);

Earlier and Later Egyptian differ from one another in the following respect: the former displays a preference for synthetic grammatical structures (suffixes for gender, number, etc.), whereas the latter shows a tendency towards analytic grammatical devices (prefixation, verbal periphrases). Moreover, Demotic and Coptic differ from previous stages with respect to the graphic system.

Alphabet

Egyptian hyeroglyphs are a complex set of graphemes, variable in number according to the period considered. They are pictographic signs representing any kind of persons, animals and objects. The two components of an Egyptian word are:

(1) A sequence of *phonograms* that may be mono-, bi- or triconsonantal (according to the number of consonantal phonemes represented). The vocalic phonemes are never graphically expressed. Bi- or triconsonantal signs are usually accompanied by the so-called 'phonetic complements', i.e. monoconsonantal phonograms which specify one or more phonemes of the bi- and triconsonantal sign, thus allowing for a more simplified reading of the intended word.

(2) A *semagram,* named *determinative* in Egyptological tradition, usually accompanies the sequence of phonograms spelling out the semantic domain of the word.

Some words are written phonologically, combining (mainly mono-)consonantal signs, many others are written with logograms (referred to as ideograms in the Egyptological literature), i.e. signs either representing the object or another entity phonologically similar in structure to the intended word (as in a *rebus),* or even an object metonymically or metaphorically connected to the intended object. There are 24 monoconsonantal signs.

Hieratic and Demotic writing systems derive from the hieroglyphic one: Hieratic is an italicization of the latter, whereas Demotic stems from the late Hieratic writing. The Coptic writing uses Greek alphabet (with seven additional signs derived from Demotic and representing sounds absent from the Greek alphabet).

Phonology and Pronunciation

The study of Egyptian phonology is made difficult by the writing system used. As noticed above, vowels are never represented in the hieroglyphic alphabet, semivocalic phonemes only seldom. Things are complicated by the fact that the phonological reality of Coptic (vocalic) characters, often invoked as a comparison for the purpose of phonetic reconstruction, is far from being clear, and the Coptic is chronologically far away from the earliest stages of Egyptian language. Therefore, the pronunciation of Egyptian words is often a matter of convention. Other heuristic criteria resorted to by egyptologists in order to reconstruct the phonological reality of Ancient Egyptian are mainly the contemporary transcriptions and renderings of Egyptian words in Akkadian (archive of el-'Amarna, fifteenth to fourteenth century BCE) and in Greek texts (starting from the fourth century BCE). Also, the Egyptian rendering of foreign words may be taken as a reconstructing criterion.

The consonantal inventory of Ancient Egyptian is given in Table 1 (adapted from Loprieno 1995;

TABLE 1   The Consonantal Phonemes of Ancient Egyptian

Consonants	Bilabial	Dental	Alveo-palatal	Palatal	Velar	Uvular	Pharyngeal	Glottal
PLOSIVE:								
Voiceless	*p*/p/ [p⁽ʰ⁾]	*t*/t/ [t⁽ʰ⁾]		*ṯ*/c/ [c⁽ʰ⁾]	*k*/k/ [k⁽ʰ⁾]	*ḳ*/q/ [q']		
Voiced	*b*/b/	*d*/d/ [t']		*ḏ*/ɟ/ [c']	*g*/g/ [k']			
FRICATIVE:								
Voiceless	*f*/f/	*s*(*ś*)/s/ [sʲ]	*š*/š/	*ḫ*/ç/		*ẖ*/χ/	*ḥ*/ħ/	*h*/h/
Voiced		*z*(*s*)/z/ [s']					*ʿ*/ʕ/	
NASAL	*m*/m/	*n*/n/						
VIBRANT		*r*/ɾ/ [ɾ]				*ꜣ*/R/		
GLIDE	*w*/w/			*j*/j/				

phonemes are given both in the common transcription system of Egyptian texts and in their supposed phonetic value).

Historically, the opposition between voiceless and voiced phonemes tends to be neutralized into the voiceless variant. The vowel inventory is fairly limited (as usual in Semitic languages), including /i/, /a/ and /u/ and their long counterparts. The phonemes /e/ and /o/ are attested in Later Egyptian.

Morphosyntax

Ancient Egyptian has inherited the preference of languages of its family for bi/triconsonantal roots. The number of consonants in a root may range from one to four, but the vast majority of roots is biconsonantal *(sn* 'brother') or triconsonantal (*ḫpr* 'become'). Entire lexical families may be formed from a given root, either by varying the internal vowels or by means of suffixation. In the transcription system, suffixes are indicated by periods preceding them. As for nominal morphology, there are suffixes marking gender (masculine and feminine) and number (singular, plural, dual [attested only in Early and Middle Egyptian]). In Table 2, the morphemes for gender and number are illustrated for the noun *sn* 'brother'.

TABLE 2 Gender and Number Markers for *sn* 'brother'

	singular	plural	dual
masculine	*sn.Ø*	*sn.w*	*sn.wj*
feminine	*sn.t*	*sn.wt*	*sn.tj*

Early Egyptian has no articles: the form *sn* may refer both to 'a brother' and to 'the brother'. Later, articles arose from the deictic *pꜣ/tꜣ/nꜣ* 'the said'.

When used attributively, adjectives usually follow the noun and agree with it in gender and number (*sn=j nfr* [brother-my beautiful] 'my beautiful brother'/ *sn.t=j nfr.t* [sister-my beautiful] 'my beautiful sister'/, etc.). When used predicatively, they precede the noun and do not agree with it (the so-called adjective verbs: *nfr sn=j* [beautiful brother-my] 'my brother is beautiful'/ *nfr sn.t=j* [beautiful sister-my] 'my sister is beautiful'/ etc.).

Personal pronouns are either independent or affixal; affixal pronouns are further divided into **dependent** and **suffixed** pronouns (see Table 3, adapted from Loprieno 1995): the former are used mainly as objects of transitive verbs, and as subject of adjectival and adverbial phrases, the latter are used to indicate: (i) the possessor in nominal construct state *(sn=j* [brother-I] 'my brother'), (ii) the subject of verbal phrases *(sḏm=j* [hear-I] 'I hear'), (iii) the prepositional complement *(n=j* 'to me (DAT)'). *Independent* pronouns, on the other hand, are used as subjects in nominal phrases *(jnk sn=k* [I brother-you] 'I'm your brother') and in verbal phrases. In both these cases, the pronoun is in focus and conveys contrastive information (as in cleft sentences).

Possessive pronouns ('mine', 'yours', etc.) are formed by a demonstrative pronoun (*pꜣj/tꜣj/nꜣj* 'this (masc./fem./plur.)' agreeing with the possessee) and a suffix pronoun agreeing with the possessor. Possessive pronouns are often used as possessive adjectives in nonofficial texts instead of the suffix pronouns (*pꜣj=j sn* [the-my [mine] brother] 'my brother' instead of *sn=j* [brother-I] 'my brother'). This strategy is the most widespread in Demotic.

TABLE 3 Egyptian Personal Pronouns

	Independent	Dependent	Suffixed
SINGULAR:			
1st person	*jnk*	-*wj* (later *twj*-)	=*j*
2nd person masculine	*ṯwt* (later *ntk*)	-*kw* (later *twk*-)	=*k*
2nd person feminine	*ṯmt* (later *ntṯ*)	-*ṯm* (later *twt*-)	=*ṯ*
3rd person masculine	*swt* (later *ntf*)	-*sw* (later *sw-* > *ef*-)	=*f*
3rd person feminine	*stt* (later *nts*)	-*sj* /-*st* (later *st-* > *es*-)	=*s*
DUAL			
1st person			=*nj*
2nd person		-*ṯnj*	=*ṯnj*
3rd person	*ntsnj*	-*snj*	=*snj*
PLURAL			
1st person	*jnn*	-*n* (later *twn*-)	=*n*
2nd person	*nttn*	-*ṯn* (later *twtn*-)	=*ṯn*
3rd person	*ntsn* (later *ntw*)	-*sn/-st* (later *st*-)	=*sn* (later =*w*)

TABLE 4   Relative Pronouns

	positive (who)	negative (who … not)
Masculine	*ntj*	*jwtj*
Feminine	*ntt*	*jwtt*
Plural	*ntj.w/nt.w*	*jwtj.w*

TABLE 5   Personal Endings of the Pseudo-Participle

	Singular	Dual	Plural
1st person	*.kj > .kw*	*.wjn*	
2nd person	*.tj*	*.twnj*	*.twnj*
3rd person masculine	*.j > .w*	*.wj*	*.w*
3rd person feminine	*.tj*	*.tj*	*.tj*

Relative pronouns are 'positive' and 'negative' (see Table 4).

The possession relation between two nouns is expressed either (i) with the so-called nominal 'construct state' (i.e. the nouns are juxtaposed with no overt marking of either the possessor or the possessee; the word order of this construction is possessee–possessor: *s3 R^c* [son Ø Ra] 'son of Ra') or (ii) with the determinative pronoun *n(j)*[masc.sg.]/ *nw*[masc.plur.] */nt*[femin. sg. and plur.] 'that of, belonging to', agreeing with the possessee: *s3.t nt R^c* [daughter DET$_{+AGR}$ Ra] 'daughter of Ra'. In the development of the Egyptian language, this determinative pronoun grammaticalized into an invariable genitive marker *n: s3.t nR^c* [daughter DET$_{-AGR}$ Ra] 'daughter of Ra'.

The verbal system of Classical Egyptian is synthetic: T-A-M morphemes *(n* past tense; *t* perfective; *w* prospective aspect and passive voice; *tw* passive voice) and suffix pronouns are attached directly to the verbal root. T-A-M morphemes, if present, precede the suffix pronouns. Some verbal roots ('weak verbs') show a reduplication of the last radical phoneme in non-finite verb forms. The imperative corresponds to the basic form of the root (with no suffix pronouns or reduplication); some verbs, however, display irregular forms of imperative (*rdi* 'to give', imp.: *im*; *ij* 'to go', imp.: *mi*). A form usually called pseudo-participle, old perfective, or stative (cf. Table 5, adapted from Loprieno 1995) is characterized by a series of suffix pronouns different from that sketched in Table 3. This form conveys perfective meaning.

Infinitives bear a Ø-morpheme in the regular verbs and a *.t* feminine ending in the class of 'weak verbs'.

Sentences are classified according to the kind of predicate; compare the following nominal (1), adjectival (2), adverbial (3) and verbal (4) sentences:

(1) *ink sn=k* ('I'm your brother')
(2) *nfr pr=k* ('your house [*pr=k*] is beautiful [*nfr*]')
(3) *s3.t=k m pr=k* ('your daughter [*s3.t=k*] is in [*m*] your house [*pr=k*]')
(4) *sdm=j it=f* ('I hear [*sdm=j*] his father [*it=f*]')

Demotic

The stage of development termed Demotic is characterized by the evolution from a synthetic to an analytic morphology. In the field of nominal morphology, for instance, the development of a definite and indefinite article from the deictic adjective *p3/t3/n3*, 'the said', pairs the progressive loss of the nominal endings (see Table 2).

	Singular definite	Plural definite	Singular indefinite	Plural indefinite
Masculine	*p3 sn*	*n3 sn(.w)*	*w^c sn*	*hjn sn(.w)*
Feminine	*t3 sn(.t)*	*n3 sn(.wt)*	*w^c.t sn(.t)*	*hjn sn(.wt)*

References

Bresciani, Edda. 2002. Nozioni elementari di grammatica demotica, con Liste grafiche e letture demotiche (a c. di Angiolo Menchetti). Pisa: ETS.

Černý, Jan and Saram Israelit Groll. 1984. A late Egyptian grammar, 3rd edition. Studia Pohl, Series Maior IV. Rome: Pontifical Biblical Institute.

Erichsen, Wilhelm. 1954. Demotisches Glossar. Kopenhagen: Ejnar Munksgaard.

Erman, Adolf. 1933. Neuägyptische Grammatik, 2nd edition. Leipzig: Wilhelm Engelmann.

Erman, Adolf, and Hermann Grapow. 1926–1953. Wörterbuch der Ägyptischen Sprache, Vols. 1–4. Berlin: Akademie Verlag.

Gardiner, Alan. 1927. Egyptian grammar, being an introduction to the study of hyeroglyphs. Oxford, UK: Oxford University Press, 3rd edition, 1957.

Helck, Wolfgang, Eberhard Otto, and Wolfhart Westendorf (eds.) 1975–1992. Lexicon der Ägyptologie (LÄ), 7 vols. Wiesbaden: Harrassowitz.

Johnson, Janet H. 1991. Thus Wrote Onchsheshonqy. An introductory grammar of Demotic, 2nd edition. Chicago, IL: Oriental Institute.

Loprieno, Antonio. 1995. Ancient Egyptian. A linguistic introduction. Cambridge, UK: Cambridge University Press.

Osing, Jürgen, 1976. Die Nominalbildung des Ägyptischen, 2 vols. Mainz: Verlag Philipp von Zabern.

Polotsky, Hans-Jakob. 1971. Collected Papers. Jerusalem: Magnes Press.

Spiegelberg, Wilhelm. 1925 Demotische Grammatik. Heidelberg: Carl Winters Universitätsbuchhandlung.

ANGIOLO MENCHETTI

See also **Afroasiatic; Coptic Egyptian**

Ancient Greek

Overview

A(ncient) G(reek) belongs to the Indo-European family, of which it constitutes a separate branch. Its primary relevance to Indo-European studies and studies on language change owes to several factors. In the first place, AG is the second earliest attested Indo-European language after Hittite. It is considered to be pretty conservative under different respects (phonology, inventory and realization of inflectional categories, word order, among others). Besides, its development continued after antiquity into the Middle Ages (Byzantine and Medieval Greek) up to the present (Modern Greek), which makes Greek the language with a longest recorded history together with Chinese. Unfortunately, this long time span is not evenly documented: there are breaks, one of which concerns the present discussion, covering the period that goes from the first written attestations (13th century BCE) to the redaction of the Homeric poems (eighth century BCE).

Speakers of some AG varieties are thought to have entered mainland Greece in the second half of the third millennium BCE, then spreading to the Aegean islands, Crete and Cyprus, and slightly later to the Aegean coast of Anatolia, and to Southern Italy and Sicily. The first written sources date back to the second millennium BCE, and were found in Pylos (in the Peloponnesus) and on the island of Crete. The language of these early records is usually referred to as Mycenean Greek.

Although they are several centuries older than the main bulk of AG written sources, the Mycenean tablets do not preserve a sort of "Proto-Greek": in spite of displaying a number of archaic features, including phonemes that later merged with others in all Greek varieties, Mycenean is already characterized as belonging to a specific dialectal group. Dialectal variation is another noteworthy peculiarity of AG: contrary to most ancient languages, AG had various local literary traditions, based on different varieties; furthermore, inscriptions preserve local vernaculars, still different from the standardized literary variety chosen for each dialect. Although variation within the literary language is mostly limited to Pre-Classic time (before the fifth century BCE), specific literary genres remained connected with the dialect in which they had their earliest development, so that classical authors used to give a specific dialectal color to certain types of texts (for example, in Attic tragedy it was customary to introduce a number of Doric features in the choruses, because the invention of choral poetry attributed to the Dorians).

The two most important literary dialects were Ionic and later Attic. Since these two varieties are closely related to each other, and literary Attic gave up a number of vernacular features under the influence of the more prestigious Ionic, the literary dialect is commonly known as "Attic-Ionic". The preservation of dialectal variation was favored by political fragmentation. After the unification of Greece under the Macedonian kings in the fourth century BCE, linguistic unification also started, resulting in the development of a common variety, the *koiné,* which was spoken during the centuries of the Roman (later Byzantine) Empire, and served as the basis for further developments.

Mycenean Greek

M(ycenean) G(reek) is known since the 1950s, when its decipherment was accomplished by Ventries and Chadwick. The Mycenean script, commonly known as Linear B, is a syllabary consisting of around 90 signs for CV syllables, which makes it unsuitable for writing consonant clusters and word final consonants. The clay tablets that document MG were not meant for long term preservation, as shown by the fact that they were not baked: some of them reached us because they were accidentally baked in the fires that destroyed the archives. The tablets mostly contain lists of various items, lists of workers, and some descriptions of land properties. In spite of the limitations connected with the text types and of poor spelling, a number of interesting features can be singled out, including the preservation of the Indo-European labiovelars, changed into bilabial or dental stops in later dialects, and of velar glides, dropped in the later literary varieties, a richer case system, which included at least an independent instrumental case, and systematic omission of the verbal augment. Other peculiarities of MG show its dialectal connections with the Arcado-Cypriot and the Aeolic groups.

The Greek dialects

The Greeks, who lacked political cohesion, identified themselves as opposed to all other peoples on the basis of what they felt as a common language. They divided themselves into three dialectal groups, Ionic (also comprising Attic), Aeolic, and Doric, all connected

with prestigious literary traditions. Modern research has shown that a fourth dialectal group also existed, commonly called Arcado-Cypriot, possibly not recognized by the Greeks because its speakers did not produce any noteworthy literary achievement; furthermore, Doric is now considered a member of a wider group, called Western Greek. Until the decipherment of MG, Greek dialects were classified as belonging to two main groups, Eastern Greek, further divided into Attic-Ionic, Aeolic, and Arcado-Cypriot, and Western Greek, comprising Doric and North Western Greek. The similarities of MG to both the Arcado-Cypriot and the Aeolic groups have led certain scholars to set up a three-way classification, including Attic-Ionic, Achaean (Aeolic, Arcado-Cypriot, and Mycenean), and Western Greek (see Figure 1).

(NB: Southern Italy and Sicily were occupied by speakers of Doric).

Numerous dialectal variants were simply phonetic, although some phonological differences also existed; besides there were morphological and lexical features peculiar of each dialectal group, but all varieties were by and large mutually intelligible.

Written sources

Apart from the Mycenean tablets, the oldest written source is constituted by the Homeric poems. The poems were first written in the eight century BCE (the oldest epigraphic inscriptions also date back to the same century), but their oral composition dates back to several centuries before. Writing had disappeared from Greece after the fall of the Mycenean civilization; early historians trying to reconstruct Greek history describe a sort of "Dark Age" with rather turbulent social habits; they had no memory of the refined civilization that preceded them, and did not know that earlier Greeks had used another writing system centuries past. The Homeric poems were composed during this illiterate time; their language, basically an Ionic variety, preserves a number of archaic features, which makes the poems one of the most important documents for Indo-European studies.

Other early literary texts include Aeolic and Doric poetry from the seventh and sixth centuries; prose was apparently an Ionic invention, the most significant early Ionian prose writer being Herodotus, who composed his *Histories* in the sixth century BCE. The bulk of what is commonly considered classical Greek literature, from the fifth and early fourth centuries BCE, comes from the city of Athens. It is written in literary Attic, and includes the tragedies of Aeschilus, Sophocles, and Euripides, Aristophane's comedies, as well as works by philosophers as Plato, historians as Thucidides, and numerous political orations.

The language of later writers increasingly displays typical features of the *koiné*. One of the most important sources for the knowledge of the latter is constituted by the *New Testament,* written in the first century CE. From the same period we also have a number of non-literary papyri, which help us understand the real extent of changes undergone by the spoken language. For during the first centuries CE literary Greek started to move more and more away from the spoken language, first with a puristic reaction to the intrusion of "vulgar" features (Atticism), and later with the development of a peculiar diglossic situation, which was still characteristic of Modern Greek until the 1960s.

AG was written in alphabetic script, with a wide number of local variants. Table 1 contains the standard Greek alphabet, as used in modern editions of ancient authors; it basically corresponds to the Modern Greek script, although phonetic values are no longer the same.
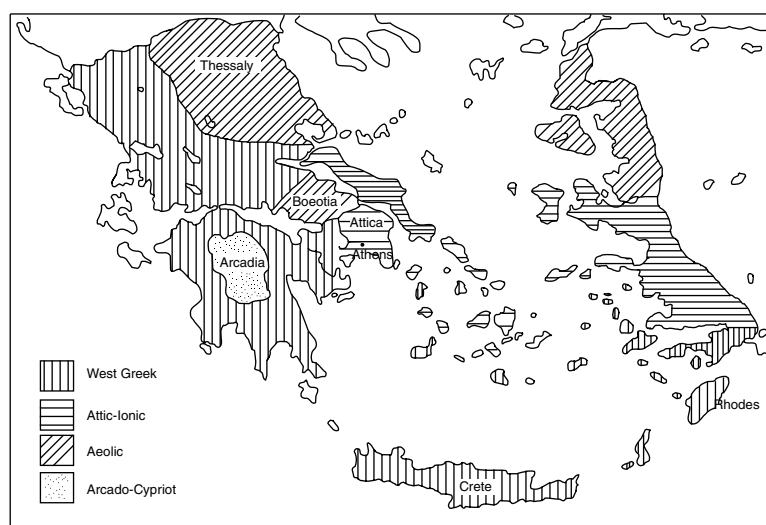
Figure 1. The Ancient Greek dialects (from Horrocks, 1997: 8).

Grammatical sketch of AG

Phonology

AG preserved the five vowels system of P(roto)-I(ndo)-E(uropean), and the distinctiveness of vowel length (long vowels are marked by ':') (see Table 2).

For ancient greek consonants, see Table 3.

TABLE 1 The Greek Alphabet (from Horrocks, 1997: xix-xx)

Greek letter		Ancient pronounciation	Modern pronounciation
*A*α	(alpha)	[a, aː]	[a]
*B*β	(beta)	[b]	[v]
*Γ*γ	(gamma)	[g]	[γ, j]
*Δ*δ	(delta)	[d]	[o]
*E*ϵ	(epsilon)	[e]	[e]
*Z*ζ	(zeta)	[zd]	[z]
*H*η	(eta)	[ɛː]	[i]
*Θ*θ	(theta)	[tʰ]	[θ]
*I*ι	(iota)	[j, iː]	[i, j]
*K*κ	(kappa)	[k]	[k]
*Λ*λ	(lambda)	[l]	[l]
*M*μ	(mu)	[m]	[m]
*N*υ	(nu)	[n]	[n]
*Ξ*ξ	(xi)	[ks]	[ks]
*O*o	(omikron)	[o]	[o]
*Π*π	(pi)	[p]	[p]
*P*ρ	(rho)	[r]	[r]
*Σ*σ/ζ	(sigma)	[s]	[s]
*T*τ	(tau)	[t]	[t]
*Y*υ	(upsilon)	[y, yː]	[i]
*Φ*φ	(phi)	[pʰ]	[f]
*X*χ	(chi)	[kʰ]	[x, ç]
*Ψ*ψ	(psi)	[ps]	[ps]
*Ω*ω	(omega)	[oː]	[o]

TABLE 2 Ancient Greek Vowels

/i/		/u/	/iː/		/uː/
	/e/	/o/		/eː/	/oː/
	/a/				/aː/

TABLE 3 Ancient Greek Consonants

voiceless stops	p	t	k	
voiced stops	b	d	g	
aspirated stops	pʰ	tʰ	kʰ	
fricative		s		h
nasals	m	n		
lateral		l		
vibrant		r		
glide	w			

Morphology

AG has a big number of inflectional categories, both for nouns and for verbs. Its morphology is highly fusional, with a fairly high degree of allomorphy, partly due to simplification of consonantal clusters or monophthongization of diphthongs (vowel contraction).

Nouns are inflected for case (nominative, accusative, genitive, dative, and vocative) and number (singular, plural, and dual); they belong to three genders (masculine, feminine, and neuter). Nominal inflection only employs suffixes. There are three inflectional classes, called first (/a:/ stems), second (/o/ stems), and third declension (consonant and /i/ and /u/ stems); assignment of a noun to a declensional class is determined by its phonological shape. Already in MG /a:/ stems, which were mostly feminine, developed a subclass for masculine nouns, with special endings for the nominative and the genitive singular. The third declension, which displayed the highest degree of allomorphy, lost its productivity rather early, and nouns of the third declension moved to other inflectional classes (some doublets with a consonant and a -o- variant are already found in Homeric Greek; in the *koiné* consonant stems switched massively to -*ā*- stems).

Adjectives, demonstrative, and the definite article (created after Homer) share the inflectional categories of nouns; when modifying nouns they agree with their heads in all inflectional categories. Adjectives are divided into two classes, according to whether they have three different forms for the three genders, or only two (with masculine and feminine undifferentiated). They also inflect for degrees of comparison (comparative and superlative).

Verbs are divided into two classes, thematic and athematic. They are inflected for mood (indicative, optative, imperative, and subjunctive), tense/aspect (present, imperfect, aorist, perfect, future, pluperfect, and future perfect), diathesis (active, middle, and passive), person, and number (singular, plural, and dual). A separate passive is only found in the aorist, perfect, and future; in the other tenses, the middle can have both middle and passive meaning; the imperfect, the perfect and the pluperfect are found in the indicative only, the future and the future perfect only in the indicative and in the subjunctive. Nominal forms of the verb include an array of infinitives and participles and two verbal nouns. Verbal morphology is more varied than nominal morphology: besides suffixes, one also finds prefixes (so-called "augment" for past tenses in the indicative, and reduplication in the perfect and some presents), and alternation of the root vowel (apophony). Below some typical verb forms are analyzed for convenience:

1. *leíp-ō,* first sg. pres. indic., "I leave"
   present stem+inflectional ending

2. *leíp-oi-mi,* first sg. pres. opt., "may I leave"
   present stem+suffix+inflectional ending
3. *é-leip-on,* first sg. imperfect, "I was leaving"
   augment+present stem+inflectional ending
4. *é-leip-s-a,* first sg. aorist indicative, "I left"
   augment+present stem+suffix+inflectional
   ending
5. *é-lip-on,* first sg. aorist indicative, "I left"
   augment+aorist stem+inflectional ending
6. *le-loíp-a,* first sg. perfect, "I left"
   reduplication+perfect stem+inflectional ending

(this verb has two possible aorists, sigmatic, as in (4), with the suffix -*s*- and the same stem as the present, and thematic, as in (5), with apophonic variation of the root vowel).

AG also has a big variety of derivational affixes. Most category changing affixes are suffixes: *díkē,* "right (noun)", *dík-aio-s,* "just (adjective)", *dik-aio-súnē,* "justice (noun)", *dik-áz-ō,* "I judge (verb)". Prefixed derivates normally maintain the same word class of the base: *a-díkaios,* "unjust (adjective)", *sun-dikázō,* "I have a share in judging (verb)". Besides derivation, compounding was also very productive.

Syntax

AG has nominative-accusative alignment. The subject of both transitive and intransitive verbs is inflected in the nominative; note, however, that neuter nouns have nominative=accusative. A further peculiarity of neuter subjects is constituted by incomplete verbal agreement: neuter plural take the verb in the singular: *tà zôia trékhei,* "the animals (pl.) run (3rd sg.)". The subject need not be overtly expressed when it is recoverable from the context; to a lesser extent, the direct object, too, may be omitted if recoverable. Transitive verbs can have passive form, in which case subject function is assigned to the direct object of the corresponding active; the agent may or may not be added as an adjunct (prepositional phrase).

The accusative is basically the case of the direct object of transitive verbs; it also has a limited use as allative (mostly in poetry), and a rather peculiar adverbial use, the "accusative of respect": *pódas* (accusative) *ōkùs* (nominative) *Akhilleús* (nominative), "Achilles swift-footed".

The genitive is used for nouns functioning as modifiers of other nouns. It also has an adverbal use: certain verbs only take the genitive (*árkheín tinos,* "to rule over sbd. (genitive)"), others allow variation of the accusative and the genitive, which in such cases has partitive meaning: *píen tòn oînon / toû oínou,* "to drink (all the) wine (accusative)/ some wine (genitive)". At a pre-literary time the genitive merged with the ablative, and took over its functions; consequently, we find the genitive with verbs that indicate some sort of separation and with the prepositions that take the ablative in the other Indo-European languages (e.g. *ek,* "out of", *apó,* "from", cf. Latin *ex, ab,* both with the ablative).

The dative is the case of the third argument of verbs of exchange and communication; when occurring outside the verbal valency, it can indicate the participant for whose sake an action is performed (benefactive); often it occurs in the "external possessor construction": *epeidḕ dé moi* (dative) *hē mḗtēr eteleútēse,* "after my mother died", "after the mother died on me". The dative merged with two other cases: the locative (possibly already before the Mycenean age), and the instrumental (after MG, where the instrumental case is still preserved). While the locatival function of the plain dative is mostly limited to poetry, its instrumental function was very productive: only around the first century CE prepositional phrases start to be found in place of the plain dative to express instrumental.

Verbal tenses partly have aspectual value, especially outside the indicative mood. The basic aspectual opposition is between the present (imperfective) and the aorist (perfective). The perfect originally had a resultative function, inherited from PIE: pres. *baínō,* "I go", perfect *bébēka:* "I am (in a certain place)". Modality is expressed partly by verbal moods, partly by the modal particle *an,* used in some conditional sentences, and partly by verbal tenses (the imperfect may be used for contrary-to-fact modality). Besides, the optative also has a purely syntactic use when it occurs in subordinate clauses that depend on other subordinate clauses ("oblique optative").

AG has a complex system of hypotaxis, including: (a) complement clauses, which can have the verb in the infinitive and the subject in the accusative or in the nominative, or the conjunction *hóti,* "that", and the same moods/tenses of independent clauses; (b) relative clauses, with the relative pronoun *hḗ, hós, hó;* (c) various types of adverbial clauses. A peculiarity of AG subordination is the extremely widespread and elaborated use of participles. The wide use of participles can at least in part be explained through the need to make up for the lack of adverbial verb forms ("converbs", corresponding to the English gerund); still, the existence of participles for most aspect/tenses and for all diatheses constitutes a rather infrequent typological feature of AG.

AG word order is governed by pragmatic factors. In particular, the finite verb can stand in any position in the sentence, and, if transitive, precede or follow the direct object, both in poetry and in prose. The subject, if expressed, can be placed before or after the verb, but preverbal position is more frequent. Modifiers (both adjectival and nominal) can precede or follow their heads; adpositions mostly precede their complements,

although Homer has some occurrences where they are postposed, and relative clauses follow their head, with few exceptions. The definite article precedes the noun or noun group it determines; if a head noun is followed by an attribute, the article is repeated (*ho anèr ho agathós,* "the good man", lit. "the man the good"). A noteworthy group of words with respect to word order is constituted by so-called "postpositives". Postpositives are various types of conjunctive particles, modal particles, and enclitic pronouns; they all share the peculiarity that they cannot be placed in sentence initial position (this opposes them to prepositives, as the negation *ou,* that can never be sentence final). Most frequently, postpositives are placed after the first word or first constituent in the sentence (Wackernagel's position), thus separating parts of the same constituents, sometimes even the definite article from the noun it determines.

Greek and Indo-European

AG is generally considered a rather conservative Indo-European language. Features that make AG particularly valuable for the reconstruction of PIE include the vowel system, diphthongs, the three series of stops (voiceless, voiced, and aspirated), the retention of numerous inflectional categories, which AG shares with Indo-Iranian, and many of the inflectional affixes, along with the different morphological processes (reduplication, root and suffix apophony, pre- and suffixation) and a highly fusional morphological technique. In spite of case syncretism, AG still preserves the basic uses of cases which can be reconstructed for PIE. Word order variation connected with pragmatic factors is common to many ancient Indo-European languages, although AG appears to have moved in the direction of bigger freedom: in this connection, the retention of Wackernagel's Law for the placement of clitics and other postpositives is a remarkable archaic feature. The field in which AG has innovated most is certainly hypotaxis, whose development is a consequence of the creation of a highly elaborated literary standard.

References

Allen, W. S. 1987. Vox Graeca. Oxford: Oxford University Press.
Blass, F., and Albert Debrunner. 1984. Grammatik des neutestamentlichen Griechisch. Göttingen: Vandenhoek & Ruprecht.
Buck, Charles Darling. 1955. The Greek dialects. Chicago: Chicago University Press.
Chantraine, Pierre. 1953. Grammaire homérique. Tome 2: Syntaxe. Paris: Klinksiek.
Chantraine, Pierre. 1961. Morphologie historique du grec. Paris: Klinksiek.
Chantraine, Pierre. 1968. Dictionnaire étymologique de la langue grecque. Paris: Klinksiek.
Horrocks, Geoffrey. 1997. Greek. A History of the Language and its Speakers. London, New York: Longman.
Meillet, Antoine. 1920. Aperçu d'une histoire de la langue grecque. Paris: Hachette.
Palmer, Leonard. 1980. The Greek Language. Atlantic Heights, NJ: Humanities Press.
Schwyzer, Eduard. 1939. Griechische Grammatik. Band 1. Munich: Beck.
Schwyzer, Eduard, and Albert Debrunner. 1950. Griechische Grammatik. Band 2. Munich: Beck.
Ventries, Michael, and John Chadwick. 1973. Documents in Mycenean Greek. Cambridge: Cambridge University Press.

Silvia Luraghi

Animals and Human Language 1: Overview

How is human language related to animal communication systems? While humans are animals with a common chimpanzee–human ancestor living as recently as five million years ago, neither chimpanzees nor any other species naturally use a communication system with anything similar to the complexity of human language. Is human language an enormously elaborated development of an original communication system used by the chimpanzee–human ancestor? Or is it a purely human innovation, unrelated historically to any animal 'language'? There is no consensus on these questions. These two views, 'continuist' and 'discontinuist', respectively, have vigorous champions. Both camps agree, however, that human language and animal communication systems differ significantly. It is these differences, as well as similarities, that provide linguists an opportunity for analysis of the relationship, if any, between human and animal 'language'.

It is tempting to think of language primarily as a tool for objective communication. However, humans also use language to think and to create a mental representation of their experience. Even a hermit who has never spoken for decades will use language for this purpose. Some 'discontinuist' scholars have argued

that the use of language for communication is historically secondary, and that its original function was cognitive. This would render the differences between human language and animal communication less surprising. On the other hand, it is hard to show conclusively that no animal 'language' has cognitive as well as communicative functions. This alleged peculiarity of human language is therefore controversial.

While one commonality among animal communications and human languages is the concern with biological, objective, or factual information, much human discourse is not primarily factual in content. Many animals vocalize to advertise territorial control, or for sexual display. Here, it is not 'what is said' that matters, but merely the fact that the vocalization occurs, and perhaps also its vigor and elaborateness. Similar in function to human singing, factual communication is the primary purpose of animal communication. Human language, however, can also concern abstract thought.

Every normal human adult human has a vocabulary of thousands of items. Among animal communication systems, the total does not exceed a few dozen. This may ensure that the domains of experience within which communication is possible among animals are severely limited, extending scarcely beyond food, danger, sex, and dominance.

The size of human vocabularies is one of the factors that allow us to talk about things outside our immediate environment. The fact that a person lives in New Zealand in the twenty-first century does not prevent her from talking about the surface of the planet Mars, or about the civilization of Egypt in the second millennium BCE. This characteristic of language has been called 'displacement'. In principle, displacement does not presuppose a large vocabulary. A chimpanzee, for example, could conceivably use its danger calls to talk about not immediate danger but a possible danger tomorrow. However, no chimpanzee does this. The only animal communication system that resembles language in respect of displacement is the dance 'language' of bees, which encodes information about the direction, distance, and quality of food sources. Until the 1980s, it was widely thought that animal calls were all subjective or attitudinal, concerned with the animal's state of body or mind ('I'm afraid!', 'I'm hungry!', 'I want sex!') rather than with identifiable external characteristics of their environment. It has now been demonstrated, however, that vervet monkeys use distinct alarm calls for distinct predators (eagles, leopards, and snakes). Furthermore, on hearing such a call a vervet reacts to the call itself, not to the predator whose presence is signaled. These calls lack displacement, but in other respects they have much of the objective semantic character of words in a human language.

A vocabulary of 10,000 items would be hard to manage if each item were an unanalyzable whole. However, the words of human languages can be divided into individual contrasting speech sounds. Approximately 45 (depending on dialect) of such sounds are used in spoken English, while other languages may differ in the number of discrete sounds. All languages share the characteristic of being structured, or patterned, on two levels: a level of meaningful units (words, phrases, sentences) and of meaningless units (sounds, syllables). For example, in the word *pot,* no one sound, *p, t,* or *o,* carries meaning. Of the word *top,* containing the same sounds, the same is true. Rather, it is the relationship of the sounds that imparts meaning to the words. Animal communication systems lack this so-called 'duality of patterning'. Individually meaningful calls are not constructed entirely out of meaningless parts recurring in other calls.

In human language, words are not merely strung together, but are combined into larger units (phrases and sentences) that are systematically interpretable. Thus, for example, all speakers of English will agree on the interpretation of *John will kill the crocodile*, and will agree in interpreting it differently from *The crocodile will kill John.* Here, it is the relationships among words that imparts meaning. This characteristic of language is called syntax. The syntax of human language is open-ended in that phrases and sentences can be embedded within larger phrases and sentences, as in *Mary thinks her father denied that John will kill the crocodile*. No naturally occurring animal communication system has a systematic syntax, even of the most rudimentary kind. In particular, no such system has the open-endedness that human language derives from embedding.

Some apes (mainly chimpanzees) that have been exposed to American Sign Language (ASL) have acquired vocabularies of a hundred or more signs, and have used them to communicate with humans and with each other. However, claims that these apes have mastered elements of ASL syntax are highly controversial. Controversial, too, is the claim that a bonobo, or pigmy chimpanzee, called Kanzi had at the age of eight acquired a systematic mastery of elements of English syntax. What apes can achieve in situations of exposure to human language is remarkable, but their achievements still differ considerably from those of humans. Perhaps, then, it is not in having language that humans are unique, but in having syntax.

While sharing some features, animal communication differs from human language in terms of size of vocabulary, contextual dependence on the environment for communication, and syntax. Examining these aspects of both types of communication systems, and the overlap between them, enables a fuller

understanding of language structure and utility. This more complete analysis could aid research in speech and language disorders and cognitive psychology.

References

Bickerton, Derek. 1990. Language and species. Chicago: University of Chicago Press.

Calvin, William H., and Derek Bickerton. 2000. Lingua ex Machina. Cambridge, MA: MIT Press.

Carstairs-McCarthy, Andrew. 1999. The origins of complex language. Oxford: Oxford University Press.

Cheney, Dorothy, and Robert M. Seyfarth. 1990. How monkeys see the world. Chicago: University of Chicago Press.

Deacon, Terrence. 1997. The symbolic species. New York: W.W. Norton and London: Allen Lane.

Hauser, Mark D. 1996. The evolution of communication. Cambridge, MA: MIT Press.

Hockett, Charles F. 1960. 'The origin of speech'. Scientific American 203(3). 88–96.

Savage-Rumbaugh, E. Sue, and Roger Lewin. 1994. Kanzi: the ape at the brink of the human mind. New York: Wiley.

ANDREW CARSTAIRS-MCCARTHY

See also **Animals and Human Language 2: Dolphins; Animals and Human Language 3: Parrots; Animals and Human Language 4: Primates; Syntax**

Animals and Human Language 2: Dolphins

Can a dolphin master language? Scientists cannot begin to answer this question until human language has been fully mapped and its defining features agreed upon. In the meantime, those interested in comparative cognition and in the evolution of human language can make great strides by asking smaller and more specific questions. Can a dolphin master an artificial signal system; that is, using signals that are not innate? Can it master a discrete–combinatorial system—here, one in which these artificial signals, each of fixed form, can be combined with each other into strings? Can it further master sequential rules, in which (functionally paralleling grammatical rules) the order of signals affects interpretation?

Since the 1950s, three groups of researchers have worked with bottle-nosed dolphins on the production or comprehension of such artificial, discrete–combinatorial systems, or ADCs. However, virtually all of what is currently known and accepted about dolphins' ADC abilities comes from the work of the last and current group, led by Louis Herman, and most of these data come from two wild-caught female dolphins, Phoenix and Ake. (As is usual in animal language research, individual animals are reported by name.) This work has shown that dolphins can comprehend a small discrete–combinatorial system involving a few sequence-based rules. Dolphin abilities appear to include the following:

(1) Both auditory and visual modality signals. Phoenix obeys computer-generated whistles, while Ake obeys gestures. Many nonhuman animals can only learn easily in one modality; dolphins, like apes and humans, are cognitively flexible enough to master more.

(2) Symbol comprehension. The dolphins' signal set, though small, is open, since the dolphins have mastered signals for objects they could never have encountered in their evolutionary history, (such as frisbees and surfboards). They extend established signals to new examples, and they can comprehend such signals even when they are given for the first time as part of a new function (e.g. a question rather than a command). They can also master displacement, the use of a signal for an object that is not present, indicating that for the dolphin, a signal 'calls to mind' the referent of that signal. All this suggests that dolphins to this extent genuinely comprehend symbols, and are not just responding instrumentally.

(3) The use of signal order. Each of the dolphins has been given a different 'grammar'. (For example, to ask Phoenix to put the small ball into the white hoop, the signal string is SMALL BALL IN WHITE HOOP, while for Ake the same command is WHITE HOOP SMALL BALL IN). Thus, Ake, at least, cannot be memorizing a signal sequence and executing it mindlessly, one signal at a time, in the same order as it is presented. Furthermore, their equal success with two different orders is additional evidence that they are indeed operating from sequence-based rules.

The use of order to assign roles is robust; the dolphins are far less prone than small children to use only the meaning of signals to make sense of strings that violate established sequential rules (e.g. TOSS LEFT FRISBEE rather than the correct LEFT FRISBEE TOSS). Furthermore, if given a 'grammatically' peculiar signal sequence, the dolphins will extract any correct subsequence and act on it, whether it is at the beginning, middle, or end of the whole string.

(4) Categorization of signals into classes (such as things, qualities, and actions). Diverse evidence suggests that, as in human language, it is these categories rather than the actual signals that are ordered in the dolphins' inner rules.

(5) The use of the sign strings in diverse functional contexts.

Most of Herman et al.'s research postdates the ape language controversies of the late 1970s, and has profited extensively from them. Partly by concentrating almost entirely on comprehension rather than production, the program seems to have eliminated most of the errors in method that discredited the early ape work, exceeds even current ape research in its methodological rigor, and is highly replicable.

However, Herman's group has been criticized, first, for not pursuing production skills. Transferring symbol use between comprehension and production can be difficult for common chimpanzees, and one small study with Phoenix suggests that it may prove difficult for dolphins. Perhaps, this is because Herman's program uses food rewards rather than communicative training to shape the dolphins' behavior, and this is a second point of criticism. In apes, giving the same response—food—to every correct signal use has been shown to inhibit development of maximally human-like symbol use.

Third, exactly how many mental rules the dolphins use to interpret their ADC strings is unknown, but (depending on how elaborate each rule is) it may be as few as two. This uncertainty is significant only if it is assumed that the goal is to produce an animal with extensive or full human-like language skills. More realistically, it gives a direction for future research; how many, and how diverse, are the sequential rules that this species can master?

Fourth, early claims that dolphins could master displacement were justly disputed on the grounds that the tests involved were of memory, rather than of ability to use a symbol to access a concept. Subsequent tests have since given better proof of displacement.

Fifth, early and theoretically significant claims of mastery of recursion were also justly disputed. The evidence in question was the ability of the dolphins to respond correctly, on the first trial, to a sequence of two command strings (such as FRISBEE TOSS, BALL TOUCH). Initially offered as evidence for underlying mental recursive rules, these strings appear to require merely iteration, a much simpler rule structure. The early claims of recursion have not been repeated, and to date no other exploration of recursion has been reported.

Sixth, Schusterman and Gisiner have challenged the Herman group's calculations of probability for commands involving goal objects. However, it appears that regardless of which calculation is used, the dolphins do reliably 'pass the test', albeit with more mediocre scores.

Finally, Herman et al. have been criticized for applying linguistic terms such as 'sentence', 'modifier', or indeed 'language' to the dolphins' use of the ADC. What may even be called ownership of linguistic terms is a perennial argument within animal language research. Herman overtly uses linguistic terms as hueristic devices, to provoke reassessment of definitions: others insist that such terms can be used only as defined within one or another established linguistic approach. The issue seems unlikely to be soon resolved.

References

Herman, Louis M. 1988. The language of animal language research: reply to Schusterman and Gisiner. Psychological Record 38. 349–62.

Herman, Louis M., D.G. Richards, and J. P. Wolz. 1984. Comprehension of sentences by bottle-nosed dolphins. Cognition 16. 129–219.

Herman, Louis M., and Paul H. Forestell. 1985. Reporting presence or absence of named objects by a language-trained dolphin. Neuroscience and Biobehavioral Reviews 9. 667–81.

Herman, Louis M., and Robert K. Uyeyama. 1999. The dolphin's grammatical competency: comments on Kako (1999). Animal Learning and Behavior 27. 18–23.

Kako, Edward. 1999. Elements of syntax in the systems of three language-trained animals. Animal Learning and Behavior 27. 1–14.

Premack, David. 1986. 'Gavagai!' or the future history of the animal language controversy. Cambridge, MA: MIT Press.

Roitblat, Herbert L., Louis M. Herman, and Paul E. Nachtigall (eds.) 1993. Language and communication: comparative perspectives. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schusterman, Ronald J., and Robert Gisiner. 1988. Artificial language comprehension in dolphins and sea lions: the essential cognitive skills. Psychological Record 38. 311–48.

Schusterman, Ronald J., J.A. Thomas, and F.G. Wood (eds.) 1986. Dolphin cognition and behavior: a comparative approach. Hillsdale, NJ: Lawrence Erlbaum Associates.

Shyan, M.R., and L.M. Herman. 1987. Determinants of recognition of gestural signs in an artificial language by Atlantic bottle-nosed dolphins *(Tursiops trunctatus)* and humans *(Homo sapiens)*. Journal of Comparative Psychology 101. 112–25.

DOROTHEA COGILL-KOEZ

See also **Animals and Human Language 1: Overview; Animals and Human Language 3: Parrots; Animals and Human Language 4: Primates**

Animals and Human Language 3: Parrots

In an attempt to determine the relationship between human language and animal communication, researchers have often studied the potential for communication between humans and animals. Beginning in the early 1960s, most studies on animal–human communication focused on great apes and a few cetacean species such as dolphins, as these animals were assumed to possess the potential for complex cognitive abilities considered as prerequisites for language development. Birds, with a few exceptions, were rarely subjects of such studies as they were presumed to be merely talented mimics. For example, experiments on pigeons using standard laboratory paradigms demonstrated capacities far inferior to those of mammals. These results were thought to represent the abilities of all birds, despite evidence suggesting that species such as jays, crows, and parrots might be capable of more impressive cognitive and communicative feats. In particular, parrots, like primates and cetaceans, have large brains, long lives, and highly social natures; couldn't parrots also have evolved complex cognitive capacities? Could proper training enable them to demonstrate language-like abilities comparable to those of nonhuman primates and cetaceans?

Research spanning more than 20 years has provided positive answers to these questions. After being trained with a technique called the model/rival (M/R) procedure, Gray parrots (*Psittacus erithacus*) learned to use aspects of English speech. In the standard M/R protocols, two humans demonstrate the labeling ability that a bird is to acquire. One human trains the second (the model/rival), i.e. presents and asks questions about the item (e.g. 'What's here?'). The trainer rewards correct identification with the item, demonstrating referential and functional use of labels, respectively, by providing a 1:1 correspondence between label and object, and modeling label use as a means to obtain the item. The second human not only is a model for a bird's responses and its rival for the trainer's attention, but also illustrates aversive consequences of errors: trainers respond to errors by scolding, and temporarily hide the object. Birds quickly acquire referential label use with such training.

This two-way communication code was then used to test the parrots' cognitive abilities. The oldest bird in the study, Alex, labels more than 50 objects, seven colors, five shapes, up to 6 quantities, and three categories (color, shape, material). He also appropriately uses 'no', 'come here', 'wanna go X', and 'want Y' (where X and Y are appropriate location or item labels). He correctly responds to questions asking him to judge category, relative size, quantity, similarity or difference in attributes, and to demonstrate label comprehension. When shown a colored wooden block, he can respond to questions of 'What color?', 'What shape?', 'What matter?', and 'What toy?' Given two objects, he can respond vocally to 'What toy?', 'How many?', 'What's same/different?', 'What color bigger/smaller?', 'What matter bigger/smaller?' If given a tray with different numbers of, for example, red and blue balls and blocks, he can respond to 'How many blue blocks?' with the appropriate quantity label. He thus not only produces labels in response to simple queries but also comprehends questions that involve multiple labels and recursion. He combines labels to identify, request, comment upon, or refuse more than 100 items, and to alter his environment. He uses a very limited form of syntax, involving sentence frames (e.g. 'I want X') and consistent application of the adjective–noun order construction. He requests objects in their absence and refuses substitutes, showing that he reacts to the dissonance between the concept encoded in his request and the proffered item. Such data suggest that his labels are representational. He semantically separates labeling from requesting. He practices on his own and builds upon previous knowledge to learn new sounds. After learning 'gray', he spontaneously produced 'grape', 'grate', 'grain', 'cane', and 'chain'. When these words were reinforced with the corresponding object, Alex quickly associated the two.

Particularly noteworthy are additional data demonstrating that Alex's labeling goes well beyond simple association of object and word. His labeling is equally accurate when he is presented with items that are related but not identical to training objects. He also transfers use of utterances learned in one context to another. Thus, after learning to state 'none' when asked 'What's different?' about two identical items, he could, without training, give the same response when asked 'What color bigger?' to two identically sized, differently colored blocks. Although Alex lacks all but a very few verbs, he nevertheless exhibits communicative capacities once presumed limited to humans or nonhuman primates. Similar studies with younger parrots indicate comparable results, demonstrating that Alex is not an exceptional Gray parrot.

Studies on Gray parrots also show that birds given training that lacks some aspect of input present in M/R protocols (reference, functionality, social interaction) fail to acquire referential English speech. If input fails to include, for example, joint attention of the trainer and parrot to the object being labeled, Gray parrots, like young children, do not learn the label. For similar reasons, Gray parrots fail to learn referential use of labels presented only through video or audiotapes. Depending upon the organization of M/R input, these parrots may exhibit a form of mutual exclusivity similar to that of young children: they replicate the stage in which a child believes not only that every object has a label, but also that each object can have only one label. For example, what is called 'dog' cannot be called 'animal'. Such research on acquisition patterns may aid our understanding of other forms of exceptional learning: learning unlikely in the normal course of development but that can occur under certain conditions. These results have implications both for teaching psychology and linguistics students to 'think outside the box' and for designing intervention programs for children who fail to acquire normal communication skills.

Although exceptionally rich, the behavior of the parrots described here is fully equivalent neither to human language nor to the vocal behavior exhibited even by young humans. Some behavior patterns of these birds, however, are comparable to those observed in humans, nonhuman primates, and cetaceans. The parrots' behavior thus demonstrates that certain levels of communication and cognitive development, and the responsible underlying neural structures, are present in birds and not limited to primates or mammals.

References

Pepperberg, Irene M. 1990. Conceptual abilities of some non-primate species, with an emphasis on an African Grey parrot (*Psittacus erithacus*). Language and intelligence in monkeys and apes, ed. by Sue T. Parker and Kathleen R. Gibson. Cambridge and New York: Cambridge University Press.

Pepperberg, Irene M. 1991. A communicative approach to animal cognition: some conceptual abilities of an African Grey parrot. Cognitive ethology: the minds of other animals, ed. by Carolyn Ristau. Hillsdale, NJ: Erlbaum Associates.

Pepperberg, Irene M. 1998a. The African Grey parrot: how cognitive processing might affect allospecific vocal learning. Animal cognition in nature, ed. by Russell P. Balda, Irene M. Pepperberg, and Alan C. Kamil. London and New York: Academic Press.

Pepperberg, Irene M. 1998b. Cognitive capacities of avian subjects. Comparative psychology: a handbook, ed. by Gary Greenberg and Maury M. Haraway. New York: Garland Press.

Pepperberg, Irene M. 1999. The Alex studies: cognitive and communicative abilities of Grey parrots. Cambridge, MA: Harvard University Press.

IRENE M. PEPPERBERG

See also **Animals and Human Language 1: Overview; Animals and Human Language 2: Dolphins; Animals and Human Language 4: Primates**

Animals and Human Language 4: Primates

Apes (here, orangutans, gorillas and chimpanzee species) are our closest animal relatives. Can they too master language? The earliest researchers to tackle this question, in the 1930s and 1950s, were the Hayes and the Kelloggs. Each team tried to raise an infant chimpanzee to understand and produce speech. Although both chimpanzees reportedly learnt to understand some spoken English, only one produced words — a mere four, and these with much difficulty and apparently with little understanding of their meaning.

The studies were useful, however, in showing that speech is physically impossible for apes. Instead, in the second chapter of ape language research (ALR), apes were taught to use manual signs, to arrange individual plastic symbols into symbol strings, or to point to or press symbols on keyboards.

In 1969, the Gardners reported the first great success with a chimpanzee named Washoe. (As is usual in animal language research, individual animals are reported by name.) They were followed through the 1970s and 1980s with many more chimpanzee projects, led by researchers such as Premack, Rumbaugh, Fouts, and Matsuzawa, and one each with a gorilla, Koko, trained by Francine Patterson, and an orangutan, Lyn Miles' Chantek.

In these programs, many apes soon learnt to produce well over a hundred signs or symbols. They also seemed to use their signs consciously, as true symbols. Many birds and mammals can learn to press certain buttons or produce certain behaviors 'instrumentally' merely for rewards, rather like pushing buttons on a food dispenser. However, the ALR apes extended old

signs to new examples of a referent, applying a sign for 'dog', for example, to a particular dog they had never seen before. Some apes could explicitly distinguish referents (e.g. the actual dog) from the signs for referents (here, the word *dog*), while many appeared to spontaneously invent new signs. Evidence like this suggested that these apes were using their signs as true symbols in a human-like way, rather than just instrumentally.

The researchers of the 1970s also believed that they had found evidence for rudimentary syntax in apes, at the same level, and with the same patterns, as children in the two-word stage of language acquisition. Some apes, such as the Rumbaugh team's Lana, could arrange much longer strings of signs into regular orders. Others, such as the Premack team's Sarah, could react appropriately to sequences that differed in sign order alone, such as RED ON GREEN vs. GREEN ON RED.

From these circumscribed achievements, many researchers of the 1970s argued broadly that apes could master language.

Criticism of the ALR accumulated, however, and towards the end of the 1970s it reached overwhelming proportions. To varying degrees, different ALR programs up to 1980 suffered from the following:

(1) *Redundancy and formulae*: Some apes could produce and respond correctly to long strings of signs, but these strings commonly took on a few formulaic patterns. Rather than manipulating grammatical rules, the apes could have been simply learning a few arbitrary sign sequences in which one or two elements could be varied depending on the context. The apes, it was argued, were merely pressing rote sequences of buttons on a machine, without seeing the buttons as symbols or the sequences as grammar-governed: it was the researchers who were assigning words to buttons and thus translating, for example, the button strings 21593 versus 21598 into 'Please machine give me M&M' versus 'Please machine give me milk'.

(2) *Contextual cues*: Training and testing of signs and strings often took place in sets of many, virtually identical repetitions. How much of the apes' success was due to grasping the signs themselves, and how much was due to their picking up the general pattern of the exercise?

(3) *Paralinguistic cues*: It was often argued that the apes could have been reacting to cues in the body language and gaze of the researchers, rather than to the signs themselves.

(4) *Overgenerous interpretation*: Researchers, it was claimed, were like doting parents, interpreting the random or the natural gestures of the apes as meaningful learnt signs. Secondly, even where clear signs were produced in appropriate contexts and with the same two-word grammar as young children, critics said that the apes' strings should not be assigned a grammar as, unlike children, the apes did not later develop a fuller grammar to justify such 'rich interpretation'.

(5) *Statistical inadequacy*: Those researchers who did undertake careful comprehension tests often gave such restricted answer choices that a high percentage of correct responses could have arisen from chance alone.

(6) *Lack of first trial data*: When tested with new sentences, apes were shown to respond quickly, However, there is a crucial difference between 'quickly' and 'on the first trial'. Only the latter can be used as evidence that the apes have access to grammatical rules, by which to decode the novel sentence: as they are known to quickly master new variations on old routines, it is possible that quick (but not instant) comprehension of novel sentences is merely quick mastery of a new variation on a rote task. The research of the 1970s typically did not identify this crucial first trial data.

(7) Indeed, methodology and testing in general were underreported in the work of this time.

The turning point, however, was the work of a single research group, led by Terrace and working with one chimpanzee, Nim. Careful analysis of certain videos of Nim's signing revealed the following:

(8) His 'grammatical' regularities could be ascribed to imitation. Nim certainly tended to order signs as his teachers did, but the videos showed that his sign strings were commonly repetitions or minor variations on those his teachers had just produced. His 'grammar' could therefore have been the result of simply parroting the true grammar of his human companions.

(9) His 'sentences' could be unusually long, but these long strings consisted of empty repetition. After entering the 'two-word stage', Nim's sign strings never increased in grammatical or semantic complexity. Children, conversely, only produce longer sentences on demonstrably acquiring further grammatical structures.

Terrace then analyzed shorter video segments of apes in other programs, and asserted that the same pattern of imitation could be seen in them.

Further, symbol use came under scrutiny once more. Were the apes consciously using symbols to refer, or were they just 'button-pushing' — behaving

instrumentally? Savage-Rumbaugh and colleagues then demonstrated that:

> (10) Their apes, which used signs appropriately in familiar contexts, were often unable to use the same signs in new tasks. If the apes were genuinely processing the signs as symbols, this failure is unexpected. It suggested that these apes were indeed simply producing signs instrumentally.

Finally, apart from all these methodological flaws, critics could always point to the enormous gap between the small vocabulary, two-word combinations and food-focused conversations of the apes, and the massive vocabulary, complex grammar, and wide-ranging topics of an adult human. For many critics, these differences alone were enough to justify the broad conclusion: 'no ape has shown any abilities relevant to language'.

Thus, in the late 1970s/early 1980s, the second chapter of ALR closed in great acrimony. Critics accused the ALR researchers of naivity and anthropomorphic bias; researchers in turn accused their critics of hysterical denial of ape–human continuity, and of constantly 'shifting the goal-posts' by hastily redefining language each time an ape passed a test of language.

Throughout the 1980s, the third chapter of ALR began. Firstly, the Nim project itself came under fire. The video segments Terrace analyzed were argued to be atypical of Nim's signing. These video segments, although long, were all taken from a single, socially and physically restrictive test situation (and one that the ape apparently found aversive). The tasks were limited to those in which the chances of instrumental signing were arguably maximized. Others who worked with this ape claimed that under better conditions, Nim signed differently.

Nim's signing was also argued to be unrepresentative of the species. The infant Nim had a chaotic social environment; in less than four years, he had a turnover of nearly 60 trainers, almost none with any prior experience of apes or signing. So stressful was this that the chimpanzee became intractable at an unusually young age. If Nim was behaving abnormally in the social context, it is possible that his 'linguistic' behavior was also abnormal.

Most importantly though, Nim, like many other ALR apes of that time, was taught signs through drills and food rewards. Thus, 'symbol' learning and use was in fact taught to these apes as an instrumental task. The natural consequence of this is that their sign use should indeed and predictably be instrumental. If so, such programs cannot reliably tell us much about an ape's ability to acquire symbols.

Savage-Rumbaugh and colleagues developed new methodologies to address this problem, finding that a more human-like social context was a key to human-like symbol use. Their two chimpanzees, Austin and Sherman, repeatedly failed to transfer sign use to new contexts. These failures were repeatedly overcome, not by yet more training for more rewards, but by adjusting the *social* interaction of the two apes from being ape-like to more human-like. Fuller, more human-like symbol use then flowed spontaneously from this.

The Savage-Rumbaugh group also opened a new chapter in ALR research with another chimpanzee, species, the bonobo. In the wild, this species exhibits, some more human-like social features than the common chimpanzee. This may be why captive bonobos appear capable of spontaneously acquiring human-like symbol use. Furthermore, they were found to acquire symbols (including comprehension of spoken words) in the same way as is natural for humans — simply by exposure as an infant to meaningful symbol use. (The latter discovery also showed that chimpanzees, like humans, have an early critical period for effortless 'language' acquisition.)

In many of Savage-Rumbaugh's critical tests, the methodological problems identified in the 1970s were eliminated. The weight of evidence is now in favor of the idea that, under appropriate conditions, bonobos and common chimpanzees can indeed acquire symbols, rather than just instrumentally manipulate signs.

Savage-Rumbaugh's team have also presented evidence that bonobos can master grammatical cues such as sign order, and even clause structure. Thus, Kanzi, their first successful bonobo, could respond appropriately and differentially to commands such as 'Put the tomato *in the microwave*' versus 'take the tomato that's *in the microwave* outdoors'; an achievement that, in children, is held to indicate a grasp of clause structure.

Nevertheless, as ALR moves into this new territory, new issues arise. Does Kanzi also exploit clause structure and symbol order to decide, for example, what tomato to take, or is his grasp of semantics so good that he is able to deduce much 'grammatical' information just from the meaning of the symbols?

Finally, brain imaging is beginning to enter ALR. At the time of writing, published reports are limited to only one common chimpanzee, from the Rumbaugh program, showing bilateral brain activity, rather than the typical left-hemisphere cortical activity of modern humans. Little can be built as yet from this solitary result, but the chapter that opens here is likely to be an exciting one.

'Can an ape learn a language?' The weight of evidence now indicates that under certain conditions some species can master symbol use. Beyond this, rather than many firm conclusions, ALR has

developed more appropriate methodologies and focused more specific and answerable research questions; 'Which aspects of human language can an ape master? Under what social and training conditions? How, within the brain, does it process its language-like behavior?' This is not surprising and should not be disappointing. In any rapidly developing field, most past research, from the perspective of the present, consists of demonstrations of what does not work and what is not true. ALR is in its infancy, it is to be hoped that the field has decades, perhaps centuries, of research ahead.

References

de Luce, Judith, and Hugh T. Wilder (eds.) 1983. Language in primates. New York: Springer-Verlag.

Gardner, R. Allen, and Beatrix T. Gardner. 1969. Teaching sign language to a chimpanzee. Science 165. 664–72.

Lyn, Heidi, and E. Sue Savage-Rumbaugh. 2000. Observational word learning in two bonobos *(Pan paniscus):* ostensive and non-ostensive contexts. Language and Communication 20(3). 255–73.

Peng, F.C. (ed.) 1978. Sign language and language acquisition in man and ape: new dimensions in comparative pedolinguistics. Boulder, CO: Westview.

Premack, David, and A.J. Premack. 1983. The mind of an ape. New York: Norton.

Ristau, Carolyn A., and Donald Robbins. 1982. Language in the Great Apes: a critical review. Advances in the study of behavior, Vol. 12, ed. by J.F. Rosenblatt, R.B. Hinde, C. Beer, and M.-C. Busnel. New York: Academic Press.

Rumbaugh, Duane M. (ed.) 1977. Language learning by a chimpanzee: the LANA Project. New York: Academic Press.

Savage-Rumbaugh, E. Sue. 1986. Ape language: from conditioned response to symbol. Oxford: Oxford University Press and New York: Columbia University Press.

Savage-Rumbaugh, E. Sue, Jeannine Murphy, Rose A. Sevcik, Karen E. Brakke, Shelly L. Williams, and Duane M. Rumbaugh. 1993. Language comprehension in ape and child. Chicago, IL: University of Chicago Press.

Sebeok, Thomas A., and R. Rosenthal (eds.) 1981. The Clever Hans Phenomenon: communication with horses, whales, apes and people. New York: New York Academy of Sciences.

Terrace, Herbert S. 1979. Nim. New York: Knopf and London: Eyre Methuen Ltd., 1980.

Thompson, Claudia R., and Russell M. Church. 1980. An explanation of the language of a chimpanzee. Science 208. 313–14.

DOROTHEA COGILL-KOEZ

See also **Acquisition; Animals and Human Language 1: Overview; Animals and Human Language 2: Dolphins; Animals and Human Language 3: Parrots**

Aphasia

Aphasia is an acquired language disorder that occurs after language skills have been completely developed, due to focal brain damage in the hemisphere that is dominant for language (usually the left hemisphere). In this definition, several other language and speech disorders are excluded. The term 'acquired' indicates that aphasia cannot be congenital. Of course, perinatal brain damage may cause speech and language problems, but these are not referred to by the term 'aphasia'. 'Language disorder' distinguishes aphasia from pure articulatory and fluency disorders, which may or may not be the result of brain damage, but are not central to the *language* faculty. The fact that aphasia is a language disorder implies that (usually) the entire language faculty is affected, that is, speaking, comprehension of spoken language, reading, and writing, because the same language system serves these four modalities. Aphasia is by definition due to brain damage. In some psychiatric diseases (e.g. schizophrenia), the production deficits may appear similar, but as there is no focal brain damage, this is not aphasia. The term

'focal' means that the cerebral damage is restricted to rather clearly circumscribed brain lesions, as opposed to *diffuse* brain damage, that is, damage spread over the entire hemisphere. This diffuse cerebral damage may cause language problems (for example, in dementia), but these are quite different, and hence distinguishable from aphasia.

Aphasia is the result of damage to the hemisphere that is dominant for language, most often the left hemisphere. The right hemisphere is involved in communication, however. Although damage in this hemisphere does not affect grammatical or semantic processing, communication may be impaired, because of problems at the discourse level, often referred to as 'pragmatic' problems. In some left-handers, damage to the right hemisphere may result in aphasia.

Typology

The symptoms of aphasia are rather diverse and vary, depending on the site and extension of the lesion. In

general, large lesions cause more severe deficits than small lesions and anterior lesions (in the frontal parts of the left hemisphere) cause an impairment in the ability to construct sentences, whereas lesions to the more posterior parts (the temporal (and parietal) lobes) cause problems in word finding. These relatively selective impairments to the different language levels can best be explained by the so-called 'classic aphasia syndromes'. These are more or less accepted in the aphasiological world, although pure syndromes do not occur regularly (estimated proportions in the entire aphasic population are between 15 and 100%). The main types are the following.

(1) Broca's Aphasia

This aphasia type has been named after Paul Broca (1824–1880), a French neurologist who was the first to mention that impairments to the faculty of speech were caused by a lesion in the left frontal lobe, more specifically to the foot of the third frontal gyrus, now known as *Broca's area* (see Figure 1). (N.B. he did *not* describe what is now called Broca's aphasia). This aphasia type is characterized by slow, elaborated speech in which mainly content words (nouns, verbs, and adjectives) are produced; function words (determiners, prepositions, pronouns) are often omitted. This speech style is called *telegraphic speech*.

In most of the speakers with Broca's aphasia, articulation is affected as well, because the brain site responsible for articulation is so close to the area that, if damaged, it causes Broca's aphasia. Comprehension of spoken language is relatively good: in conversations, speakers with Broca's aphasia understand what is said. When tested for the comprehension of complex syntactic structures, the performance drops dramatically. It is therefore assumed that in Broca's aphasia the grammatical component of language is affected.
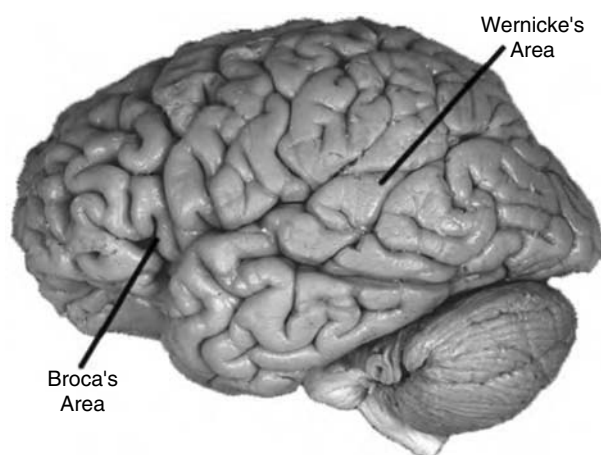
Figure 1. Left hemisphere of the human brain.

Reading and writing are also impaired, comparable to oral comprehension and production (see Figure 1).

(2) Wernicke's Aphasia

This aphasia type has been named after Carl Wernicke (1848–1905), a German neurologist who was the first to describe this syndrome. He located the lesion which causes this aphasia type in the posterior parts of the left temporal lobe (see Figure 1). Speech production is fluent and well-articulated, but the aphasic speaker produces sound and word substitutions, so-called 'paraphasias', for example 'spool' for 'spoon' or 'mother' for 'wife'. Distortions of words may be further away from the target words than in these examples, and as a consequence the listener may not recognize existing words; these sound strings are called 'neologisms', for example, 'stroofel'. When many of these words are produced, the text will no longer be understandable, which is referred to as 'jargon' or 'word salad'.

Comprehension of language is moderately to severely impaired. The combination of a severe comprehension and production disorder results in poor communication abilities. Writing and reading are parallelly affected.

(3) Anomic Aphasia

There is no clear lesion site in the brain that is responsible for this aphasia type, but the lesions are usually found in the temporal and/or parietal lobe. Anomic aphasia is characterized by word-finding problems. The oral output of a speaker with anomic aphasia is basically fluent, but due to the pauses that arise from the word-finding problems, speech rate may be low. The word-finding problems are not only recognizable by pauses, but also by the use of less specific words, such as 'thing', 'do', 'here', 'there'. The latter is called 'empty speech'. Comprehension of spoken language is relatively good, although complex commands are difficult to understand. Reading is usually spared as well, but in writing the word-finding problems show up again.

(4) Conduction Aphasia

Conduction aphasia was first described by, again, Carl Wernicke. The name of the syndrome derives from the fact that it is caused by a lesion in the tract that connects Wernicke's and Broca's area. It is characterized by the production of sound errors, for example, 'scrample' for 'scrabble'. The patient is well aware of his errors and this results in many self-corrections, for example, 'scramble, no strapple, no stample …'. The errors are particularly prominent when the patient is asked to repeat longer words and sentences. Comprehension of spoken and written language is relatively good, but

impaired for complex and long sentences. In writing, a similar error pattern is shown as in speaking.

(5) Global Aphasia

The lesion responsible for global aphasia is often large and extends to both the frontal, temporal, and parietal lobes. All language modalities are severely impaired. Spoken output is often restricted to one- or two-word utterances, sometimes only 'yes' and 'no', not always used correctly. Comprehension is severely impaired and reading and writing are hardly possible. This, of course, results in very poor communication abilities.

Again, it should be stressed that although this classification has contributed considerably to the understanding of aphasia, most patients suffer from a 'mixed' form of aphasia, which cannot be labeled with any of the types mentioned above.

Prognosis

Aphasia is caused by brain damage and this is normally irreversible. This does not mean that no improvement is possible. During the first few months and especially the first few weeks following brain damage, spontaneous improvement, if not recovery, is certainly possible. If, however, the symptoms are persistent and are still present after 3–6 months, improvement is questionable. In most countries, aphasic patients are treated by speech therapists. The goal of speech therapy is to teach the patient to deal with his language problems, through improvement of communicative strategies as well as language competence.

Psychosocial Consequences

Above, the linguistic aspects of aphasia were emphasized, since aphasia is a language disorder. One should realize, however, that a language problem has severe consequences for communicative abilities and therefore has a severe impact on the daily life of the patient and his surroundings. Nowadays, part of the speech therapy is focused on this psychosocial aspect of aphasia.

References

Overviews

Albyn Davis, G. 2000. Aphasiology: disorders and clinical practice. Needham Heights: Allyn & Bacon.

Nadeau, S.E., L.J. Gonzalez Rothi, and B. Crosson. 2000. Aphasia and language: theory to practice. New York: The Guilford Press.

Sarno, M.T. (ed.) 1991. Acquired aphasia. San Diego: Academic Press.

Visch-Brink, E.G., and R. Bastiaanse (eds.) 1998. Linguistic levels in aphasia. San Diego: Singular Publishing Group.

Clinical Aspects

Code, C., and D. Müller (eds.) 1995. Treatment of aphasia: from theory to practice. London: Whurr Publishers.

Goodglass, H., and E. Kaplan. 1983. The assessment of aphasia and related disorders. Philadelphia: Lea & Febiger.

Helm-Estabrooks, N., and A. Holland (eds.) 1998. Approaches to the treatment of aphasia. San Diego: Singular Publishing Group.

Linguistic Aspects

Grammar

Bastiaanse, R., and Y. Grodzinsky (eds.) 2000. Grammatical disorders in aphasia: a neurolinguistic perspective. London: Whurr Publishers.

Grodzinsky, Y. 1998. Theoretical perspectives on language deficits. Cambridge, MA: MIT Press.

Lexical semantics

Ellis, A.W., and A. W. Young. 1996. Human cognitive neuropsychology. Hove: Psychology Press.

Phonology

Whitaker, H.A. (ed.) 1988. Phonological processes and brain mechanisms. New York: Springer-Verlag.

Pragmatics

Joanette, Y. and H. Brownell (eds.) 1989. Discourse ability and brain damage. New York: Springer-Verlag.

ROELIEN BASTIAANSE

See also **Brain Organization and Auditory Pathways**

Applied Linguistics: Overview

Applied Linguistics, in its present form, is an emerging multidiscipline, or trans-discipline, developed on a multidisciplinary basis of inquiry into language in use, more specifically, into verbal communication within a given social and/or institutional context. Theoretical and methodological problem-oriented

approaches fundamentally define applied research objects and require the development of multidisciplinary models and methods to fit the process-oriented (rather than product-oriented) and multifaceted contextualized uses of language that researchers in this field of study are facing.

Such a multidisciplinary basis of inquiry involves a combination of approaches from linguistics and several other disciplines within the domain of human sciences, such as psychology, sociology, pedagogy, anthropology, ethnography, philosophy, politics, and psychoanalysis, as well as newer areas of investigation such as, for instance, computer sciences and cybernetics. This inquiry leads to a number of new insights and issues that are extremely relevant in response to the challenge of penetrating deeper into epistemological and ethical questions imposed by the complexity of the real world of human interactions, as well as by the dynamism of language in use, i.e. of language as *parole*, in Saussurean terms, or *performance*, in Chomskyan terms.

Thus, in applied research, real human interactions are seen as a complex object of inquiry with a crucial social and institutional, as well as a psycholinguistic dimension, and not simply as an individual or speaker/hearer subject matter. According to this view, there are also a number of political and ideological aspects of language and discourse to be accounted for, since they shape and are shaped by everyday current linguistic interactions. Similarly, in applied research, language in use is viewed as a complex processual object composed of observed units compiled into a system, which is characterized not simply by a static range of appropriate rules to constrain the shape of words, sentences and texts, but by continuous change and variability.

Although this growing research perspective began to emerge in the 1980s and covers a wide range of issues that go beyond traditional linguistic concerns, it has not rendered irrelevant the significant and ongoing role of the theoretical and methodological contributions of linguistics in applied studies. Nevertheless, by adopting a view of language as a complex and dynamic object, applied studies have challenged various aspects of linguistic tradition and have attempted to offer alternative approaches and issues. A classic example of such a result of applied inquiry is the relevant theoretical and methodological discussions and findings that led to and emerged from the sociolinguistic notion of 'communicative competence' as the ability to use language appropriately in varying social contexts, i.e. in communicative practices that are culturally and historically situated.

As many different authors have pointed out over the last two decades, through multidisciplinary research efforts, Applied Linguistics is also arguing for and responding to critical accounts of the received theories of knowledge that served as the basis of much of the work done in the field since its beginnings, in the 1950s, as a branch of linguistics. In fact, as an emerging scientific area in its own right, Applied Linguistics is increasingly interested in research questions that differ qualitatively from the language-related research carried out in linguistics, as well as from the traditional nontheoretical practical applications of mainstream linguistic findings, which developed when the term 'applied linguistics' was used to designate a subfield of inquiry within the mainstream field of Linguistics. The primary concern of early applied linguists was to develop empirical material for foreign language teaching, which was inspired and oriented by theories, methods and purposes, and, based on linguistic studies, mainly structural linguistics.

Despite the fact that this traditional role assigned to applied linguists is still emphasized by several scholars working on language learning methods and models, the fundamental reevaluation of traditional linguistic conceptions of language, subject, and context by different areas of the social sciences has begun to change the understanding of these ideas and seriously challenges the predominant modes and forms of knowledge of applied research tradition. Over the past two decades, much of the relevant discussions about epistemological and ethical issues within the subfields of bilingualism, language planning, language teaching and learning, and second and foreign language acquisition, for example, involve critical rethinking of traditional approaches.

A good example of this increasingly changing process in action in the applied research tradition is the well-established subfield of second and foreign language teaching and learning. A large body of recent work seeks to understand the role of teachers' and learners' reasons and accounts in the teaching and learning process, as well as the role of the historical, social, and politicoideological aspects of the macrosocial context of society as a whole in the microinteractional context of the school as an institution, which also shapes teacher/learner exchanges in the classroom. Another relevant contribution to this subfield has to do with language and discourse as important sites of subjectivity construction on the one hand, and as important moves in the reproduction of power relationships on the other. Indeed, the purpose of much of the contemporary research on teacher training and instructional materials and methods focuses on exploring these new concerns. Research that explores ethnographic insights and methodologies has been crucial for most of the work involving classroom linguistic interactions, as well as that of collaborative researcher/teacher work in teacher training projects.

In general terms, a great deal of recent rethinking about traditional modes and forms of knowledge in

applied research has drawn on a critical view of positivist epistemologies inherited from mainstream scientific tradition in order to face the new challenges outlined above. In response to the deficiencies of positivist models of understanding language, subject, and context, exploratory analyses of the complementary roles ascribed by other epistemologies to inductive and abductive thinking and to qualitative and quantitative data and methods, for instance, ought to be essential. The attempts to model the dynamism and variations of language as an open and changeable system have, likewise, been of fundamental importance.

References

Adger, Carolyn and Jeffrey Connor-Linton. 1997. Ethical issues for applying linguistics: introduction and prologue. Issues in applied linguistics 4(2). 169–77.

Cristal, David. 1981. Directions in applied linguistics. London: Academic Press.

Graddol, David (ed.) 2001. Applied linguistics for the 21st century. AILA Review, Vol. 14. Association Internationale de Linguistique Appliquée.

Hymes, Dell. 1972. On communicative competence. Sociolinguistics, ed. by J.B. Pride and J. Holmes. Harmondsworth: Penguin Books.

Hymes, Dell. 1992. The concept of communicative competence revisited. Thirty years of linguistic evolution, ed. by M. Putz. Philadelphia: Benjamins.

Kaplan, Robert B. (ed.) 1980. On the scope of applied linguistics. Rowley, MA: Newbury House.

Larsen-Freeman, Diane. 1991. Second language acquisition research: staking out the territory. TESOL Quarterly 25. 315–50.

Mauranen, Anna and Kari Sajavaara (eds.) 1995. Applied linguistics across disciplines. AILA Review, Vol. 12, Oxford: Catchline/AILA.

Pennycook, Alastair. 2001. Critical applied linguistics: a critical introduction. Mahwah, NJ: Lawrence Erlbaum.

INÊS SIGNORINI

See also **Chomsky; Noam; Saussure; Ferdinand de**

Arabic

Arabic is spoken natively by approximately 170 million people (Arabs) living in dozens of countries, most in what is known as 'the Arab world', which stretches from Iraq in Southwest Asia to Morocco in Northwest Africa. It is the language of prayer and scripture for approximately one billion Muslims. Originally the language of some tribes of the Arabian Peninsula, it came to be a major world language as a result of the Islamic conquests that began in the seventh century CE. In 1974, Arabic became one of the six official languages of the United Nations.

Arabic is a Semitic language, one branch of the Afroasiatic language family. Scholars long considered Arabic as a relatively pure and unchanged ancient Semitic language, practically identical to a proposed Proto-Semitic, but this position is now widely viewed as a romantic notion that saw the deserts of Arabia as the guardian of a relatively untouched language. Arabic is conservative in many respects, but also innovative. It alone preserves the full ancient Semitic case-marking system found in Old Akkadian. It participated in or was considerably influenced by some developments found only in South Semitic languages such as South Arabian and Ethiopic, as well as some found only in the Northwest Semitic language complex, which includes Aramaic and Hebrew. Arabic also exhibits innovations not found in other Semitic languages, such as a definite article with *l*.

The origins of the term "Arab" are unknown. Assyrian, Babylonian, and Hebrew use of the word refers to nomads coming from the desert. The earliest clearly attested use of Arabic (written in the Nabataean script) dates to a 328 CE tomb inscription in the Syrian desert. The oldest samples of writing (graffiti and inscriptions) in a script that can be seen as Arabic are also found in Syria and date to the fourth and sixth centuries.

The Arabic alphabet developed from that of the Nabataeans. Unlike Nabataean, Arabic is a cursive script. It consists of 28 letters and is written from right to left. As is typical of Semitic alphabets, the Arabic alphabet contains no letters to indicate short vowels. For example, a word written as *qbl* could be read in a number of ways, including *qabila* 'he accepted' or *qabla* 'before'. Like Nabataean, early forms of the Arabic alphabet failed to distinguish between some phonemes. Typically, a letter could be read in one of two ways, for example, /z/ or /r/. Given the absence of short vowels and the consonantal ambiguity, deciphering the meaning of a given text could pose problems. With the need for more precision, a system that did away with this ambiguity became standard. For example, /z/ is marked

with a dot above the letter in order to distinguish it from /r/. A method for marking short vowels and other phonological details was also devised, but these markings are used mostly in scripture and in materials for beginners, such as children's stories. In practice, the absence of short vowels in most texts is not a problem; context almost always indicates the correct reading (and where this is not the case, a marker is sometimes supplied to assist the reader). That the Arabic alphabet is relatively easy to learn comes as a surprise to foreigners accustomed to seeing the ornate and often complicated calligraphic styles of Arabic that are used in decorating mosques or book covers. Calligraphy plays a central role in Islamic art due to the general avoidance of human and animal figures in art and due to the prominent role of scripture in Islam.

The domination of consonants in Arabic orthography is indicative of a defining characteristic of Semitic languages known as the root and pattern system. Most words are based on a root consisting of three consonants. Arabic, more than any other Semitic language, excels in exploiting this morphological resource. Each root is associated with a core meaning. The systematic manipulation of these roots is the basis of Semitic derivational morphology. For example, the root *d-r-s* is associated with studying. The verb *darasa* 'he studied' is based on the template or pattern $C_1aC_2aC_3a$, the final vowel being the third-person masculine singular perfect conjugation marker; the corresponding $yaC_1C_2uC_3u$ yields *yadrusu* 'he studies' (the prefix *ya* marks it as third-person masculine imperfect and the -*u* suffix indicates indicative mood). Doubling the middle consonant (gemination) of the root—one method for forming causative verbs—produces the verb pattern $C_1aC_2C_2aC_3a$ and results in *darrasa* 'he taught'. The corresponding imperfect, based on the pattern $yuC_1aC_2C_2iC_3u$, is *yudarrisu* 'he teaches'. The corresponding active participle, based on the imperfect verb stem, is *mudarris* 'teacher/teaching' ($muC_1aC_2C_2iC_3$). The 'noun of place' pattern $maC_1C_2aC_3a(t)$—the *t* is not pronounced in some contexts—yields *madrasa(t)* 'school'.

'Arabic' is in fact a cover term for a number of language varieties that could easily be designated distinct languages, given the right social and political circumstances. The Arab world is united by a standardized variety of Arabic used mainly for writing but also for some formal oral communication. Regional differences exist but this is mostly restricted to usage and pronunciation. Linguists often refer to this officially sanctioned variety as Modern Standard Arabic (MSA). MSA is typically learned at school and most Arabs, even those who obtain advanced degrees in an Arab educational system, lack confidence (and typically experience) in using it. In contrast, Arabs grow up speaking a local variety of Arabic that differs considerably from MSA and that is officially denounced but nevertheless thrives. For example, it is the vehicle of most popular music and television programs.

MSA is the modern realization of at least the formal register of Arabic used in parts of the Arabian Peninsula in the sixth and seventh centuries. This variety of Old Arabic was apparently the language of an extensive pre-Islamic oral poetry and it is the language of the Qur'an ('reading, recitation'), the book of scripture that Muslims believe the Prophet Muhammad received over a 23-year period, beginning in 610 CE. Due to a power vacuum, within a hundred years of Muhammad's death, a relatively small number of Arabs and their allies came to control a vast empire stretching from present-day Spain and North Africa to the borders of India. In time, the religion of Islam (and, therefore, Arabic for religious purposes) spread to such far-flung places as China and what is now Indonesia (the latter being the most populous Muslim country).

The language of the Qur'an is commonly held by Muslims to be that of God. Its sound and content are felt to be so interwoven that translation is considered impossible. As a result, there are no official translations; however, translations that attempt to impart its meaning are available. All Muslims are expected to recite scripture and pray in Arabic even if they do not understand what they are saying or reading.

The nature of spoken Arabic before the Islamic conquests is debated. Early sources mention dialectal variation in the Arabian Peninsula but give only glimpses of it. Personal letters and other documents dating to the early Islamic centuries provide much more insight into the nature of early spoken Arabic. From these, it is clear that the essential characteristics of the modern Arabic dialects (collectively called New Arabic), such as the absence of case marking in nouns and adjectives, go back to at least the first centuries of the Islamic era. In other words, the language situation of the modern Arab world—where a variety of Old Arabic is used for formal purposes and a variety of New Arabic (or something akin to it) is the common language—is over a thousand years old. There is no evidence that New Arabic developed later than Old Arabic. On the other hand, there is some evidence that points to Old Arabic containing innovations not known in New Arabic or other Semitic languages.

New Arabic was possibly formed or at the very least significantly affected by the language contact situation that arose from the Islamic conquests. Arabic speakers mixed and intermarried with non-Arabic speakers and Arabic began to gain hold in places far from the Arabian Peninsula. In a gradual process that took centuries, various populations came to adopt Arabic and Islam. Subsequent migrations from the

Arabian Peninsula aided in the spread of Arabic. In the case of places like Persia (modern Iran) and Turkey, the populations embraced Islam but the relatively few Arabs who settled there eventually assimilated the language of the local inhabitants. Nevertheless, both Persian and Turkish borrowed a great deal from Arabic, including its alphabet. This is true of other Asian and African languages.

Although the pre-Islamic Arabs had a highly developed poetic tradition, they did not have a prose or academic literary tradition. This changed soon after the advent of Islam. Classical Arabic developed in response to the new demands placed on Old Arabic. It is theoretically based on the language of the Qur'an. In fact, there are distinct differences (orthographic, grammatical, stylistic…). The vocabulary of Classical Arabic expanded considerably to meet the needs of the empire by deriving new terms from Arabic roots. Classical Arabic soon became the vehicle of an impressive corpus of literary and scientific works, spurred on by the Arabs' contacts with the advanced cultures that they had conquered. For example, the Abbasid dynasty in Baghdad aggressively pursued the translation of Greek philosophical and scientific texts into Arabic. Scholars writing mostly in Arabic proceeded to considerably build on these works. These scholarly works, in turn, were passed on to Western Europe mostly via Spain and Sicily and significantly contributed to European cultural and scientific development. Particularly in the 12th century, and for some time after, learning Arabic was popular in Europe in order to read the scientific works of the Greeks (many of which were preserved only in Arabic) as well as to read the contributions of the Arabs themselves. Through such contact, European languages came to borrow numerous words of Arabic origin, such as *alcohol, asthma,* and *zenith.*

MSA is the modern manifestation of Classical Arabic. However, Classical Arabic's ornate styles and some of its complex grammatical constructions are rarely used. Again, a major difference is found in MSA's new vocabulary and phrasing. Like Classical Arabic, MSA arose in response to a language contact situation—with modern Western Europe, which is usually dated to the French invasion of Egypt in 1798. The importation of Western institutions, like newspapers, has had a profound impact on Arabic. Whereas documents composed before the modern era in Classical Arabic typically targeted a limited and elite audience, the spread of education—particularly in the second half of the twentieth century—has resulted in MSA, a version of Old Arabic streamlined for mass consumption.

The spread of education (among other factors) has also resulted in a blurring of the lines between New and Old Arabic. While mixing has long been a feature of Arabic speech communities, it is now taking place on a far wider scale. There are few 'pure dialect' speakers. MSA elements (mostly lexical) are common even in informal speech. The importation of more informal rhetorical styles (perhaps connected with Western-style democratic values or populist rhetorical approaches) may have contributed to New Arabic spreading to domains and functions formerly associated with Old Arabic. Whatever the causes, new Arabic is commonly used—with a substantial admixture of Old Arabic lexical material in particular—in contexts where it would not have been expected in the past, such as in sermons and university lectures.

A distinctive characteristic of the phonology of nearly all varieties of Arabic is the presence of emphatic consonants paired with nonemphatic counterparts, for example: /t/ and /ṭ/, /d/ and /ḍ/, /s/ and /ṣ/. Emphatic consonants are formed by raising the back of the tongue, which simultaneously results in the tip of the tongue being drawn back somewhat from the point of articulation of the corresponding nonemphatic consonant. Old Arabic has only three vowels, /a/, /i/, and /u/—as in the case of consonants, these may be lengthened with a corresponding difference in meaning. There is considerable allophonic variation in the vowels, which is conditioned by the consonants in a word. Pronunciation and word stress in MSA are typically influenced to a great degree by one's native dialect. New Arabic realizations of the phoneme corresponding to Old Arabic /q/ have become a special regional and social marker. The extreme case is Jordan, where in some rural dialects it is realized as [k], but for Bedouins it is [g]. Its urban distribution is complicated: men not of Palestinian extraction typically use [g], perhaps for its association with Bedouin toughness, whereas men from Palestinian urban families might use [ʔ], unless they want to downplay their Palestinian origins. Urban and increasingly nonurban women favor [ʔ], which is seen as sophisticated.

Much of the morphology of Arabic is a model of logic and order. This is not true, however, of the relationship between singular and plural forms of nouns and adjectives. There are two types of plurals, 'sound' and 'broken' (which designate whether the stem is left intact or is modified to form the plural). Sound plurals are formed by suffixation, for example: *mudarris/ mudarrisuuna* 'teacher/teachers'. The plural of most nouns is of the broken type and numerous patterns are used, for example: *safiir/sufaraaʔ* 'ambassador/s', *madiina(t)/mudun* 'cities', *jariida(t)/jaraaʔid* 'newspaper/s'. Some plural patterns are predictable from the pattern of the singular, most are not. Some nouns have two or more plurals.

The basic word order of Old Arabic is considered to be Verb–Subject–Object, whereas New Arabic is

classified as being of the Subject–Verb–Object type. Both, however, exhibit a good deal of flexibility. The head noun of a noun phrase precedes adjectival modifiers, which agree with the head in number, gender, and definiteness (as well as grammatical case in Old Arabic). Old and New Arabic exhibit intricate patterns of agreement variation. Numeral constructions and the negation of verbs are particulary complex in Old Arabic.

Arabic has played an important role in Western linguistics on a number of fronts. The root-and-pattern system has led to new approaches to morphology. Arabic has played a particularly prominent role in sociolinguistics. Charles Ferguson, a founding figure in the field, was profoundly influenced by his contact with Arabic. Some mark the publication of his paper 'Diglossia' in 1959 as the beginning of sociolinguistics, in which he undertook to describe the particular language situation typical of Arabic speech communities in which two closely related but substantially different varieties of a language (a 'High' and a 'Low' variety) exist side by side in a single speech community in a stable and functionally complementary relationship. This article has resulted in some 3,000 studies of the topic.

The dynamics of Arabic speech communities have also played an important role in language attitude studies, and in challenging theories of language and gender. The fact that some modern spoken varieties of Arabic are mutually unintelligible underscores the difficulty of defining precisely what a language is (as opposed to a dialect). For example, Maltese, the official language of the island nation of Malta located near Sicily, is considered by scholars to be a dialect of Arabic, whereas its speakers typically deny this relationship, prefering to see it as a Semitic language whose roots go back to ancient Phoenician. Maltese is the only variety of New Arabic to become a national language. It is written with Latin characters, with some modifications.

Arabic provides a fascinating case study in language standardization and adaptation. Few languages provide 1,400 years of documented continuous use. In many ways, it has changed remarkably little over the centuries. This applies to both Old and New Arabic, which share far more in common than is genuinely admitted. The modern dialects are so conservative that in at least some ways they are more like Pre-Islamic Arabic and the language of the Qur'an than is the standardized MSA.

The differences between MSA and an Arab's day-to-day speech are real and do present challenges. Some have called for teaching Arab children basic literacy skills in their native dialect, but there is little chance of this idea taking hold in the Arab world in the near future. Old Arabic holds tremendous symbolic value. It binds Arabs to their golden age, when they were cul-turally and scientifically superior to the West. It is a symbol of Muslim unity, as well as the cornerstone of Arab nationalism. As such, it is jealously guarded.

The realities of the modern world undercut this linguistic idealism. In their professional lives, many educated Arabs are more adept or at least more comfortable in expressing themselves in English or French than in MSA. Some influential members of Arab society criticize the Arabic abilities of recent high school graduates while simultaneously sending their own children to English or French private schools. Economic forces similar to those that resulted in their ancestors' learning and eventually adopting Arabic are now drawing young Arabs to write French email messages to Arab friends or to study medicine in English at Arab universities. In short, bilingualism is widespread and growing.

The Arab world is, nevertheless, far from being on the verge of abandoning MSA. In many ways, the modernization of Arabic (its adaptation to the needs of a changing world) over the past two centuries has been quite successful. Moreover, a far greater percentage of Arabs are literate in Arabic than at any time in history. Nevertheless, one wonders what the future holds. Arabic changed a great deal as a result of the contact situation that arose with the Islamic conquests. The present is also a time of intensive language contact. In addition to contact with the West, the very process of educating the masses has resulted in a new contact situation. Millions of Arabs with limited proficiency in MSA are using it on a day-to-day basis. While it is not likely that they will wrench the language from the hands of the traditional gatekeepers, it nevertheless seems unlikely that it could acquire so many users without their leaving their mark on it.

References

Bakalla, Mohammad H. (ed.) 1983. Arabic linguistics: an introduction and bibliography. London: Mansell Publishing.

Belnap, R. Kirk, and Niloofar Haeri (eds.) 1997. Structuralist studies in Arabic linguistics: Charles A. Ferguson's papers, 1954–1994. Leiden: Brill.

Ferguson, Charles A. 1959. Diglossia. Word 15: 325–40.

Fischer, Wolfdietrich, and Otto Jastrow. 1980. Handbuch der arabischen Dialekte. Wiesbaden: Otto Harrassowitz.

Haeri, Niloofar. 1996. The sociolinguistic market of Cairo: gender, class, and education. London/New York: Kegan Paul.

Holes, Clive. 1995. Modern Arabic: structures, functions, and varieties. London/New York: Longman.

Versteegh, Kees. 1997. The Arabic language. New York: Columbia University Press and Edinburgh: Edinburgh University Press.

R. Kirk Belnap

See also **Afroasiatic; Algeria; Diglossia; Ferguson, Charles Albert; Language: Contact—Overview; Semitic Languages**

Arabic Traditional Grammar

The essential quality of Arabic linguistics is its pragmatism, that is to say, its concern with what can be done with words in terms of the speaker's intentions, the listener's expectations, and the context of the utterance. The Arabic word for grammar itself is *naḥw*, lit. 'way, manner', i.e. way of speaking, using the same metaphor as other key Islamic terms for kinds of behavior, such as *sunna* 'way [of orthodoxy]', *šarī'a* 'path [of the law]', *ṭarīqa* 'way [of the mystic]', *maḏhab* 'way [of thinking]', i.e. 'school of thought', and it is consequently not surprising that Arabic grammatical analysis should deal primarily with the external features of language, nowadays generally called the surface structure, rather than the implicit or hidden features, viz. the underlying or deep structure. 'Meaning', therefore, is literally' what you have in mind' (*ma'nā*) or 'what you intend' (*murād*), as manifested in and only accessible through what you actually say (*kalām* 'speech').

Accordingly, the criteria for correct speech are the same as those for ethically correct behavior. A wholly successful utterance must be both 'good' (*ḥasan*, lit. 'beautiful'), i.e. well-formed, consisting of the proper words in the proper order, and semantically 'right' (*mustaqīm*, lit. 'morally straight'), i.e. conveying to the listener exactly the meaning intended by the speaker. Ill-formed utterances (*qabīḥ*, lit. 'ugly,' hence 'ethically bad') may still be 'right' if they succeed in conveying the speaker's intended meaning, just as a well-formed utterance may fail to do so, and will thus be termed *ġayr mustaqīm* 'not right'. Finally, some utterances convey no meaning at all, because of inner structural and semantic contradictions: these are labeled *muḥāl* ('wrong', lit. 'perverted, twisted'), i.e. incoherent. Significantly, in the earliest grammar at least, before the influence of Greek logic appeared, the truth or falsehood of an utterance were linguistically irrelevant.

For convenience, Arabic linguistic theory will be dealt with in three terminological groups: (1) systematic, (2) categorical, and (3) functional.

(1) The systematic terminology includes the criteria of correctness mentioned above, to which two further terms of similar ethical origin can be added: each item in an utterance has a given 'status' (*manzila*) and a corresponding 'place' (*mawḍi'*). The formal correctness of an utterance depends on using words in the place

(we would now say 'function') appropriate to their status, just as the moral value of an action is determined by the place and time of its occurrence.

It is a principle of Arabic grammatical analysis that every element operates (or not, according to its status) on the following element, this operation being termed *'amal*, lit. 'effect.' Although this very much resembles the western notion of 'government,' there is no historical or technical connection; 'government' is hierarchical, based on a vertical (paradigmatic) relationship between words, while the Arabic *'amal* is purely linear (syntagmatic), denoting the immediate effect of an 'operator' (*'āmil*) on another element 'operated on' (*ma'mūl fīhi*) sequentially.

Arabic linguistics also recognizes analogy (*qiyās*), both as the inductive device for extracting general principles from the data and as the means by which speakers create new utterances by extrapolating from speech patterns already known.

(2) Morphological categories are very few, being effectively limited to three classes of words: Noun, Verb, and Particle. This division is based on a combination of formal and semantic properties: the Noun (*ism*, lit. 'name') and Verb (*fī'l*, lit. 'act, action') are distinguished by their meaning and by certain unique features (only Nouns have Definite Articles, only Verbs have Tense, for example), while the Particle (*ḥarf*, lit. 'bit, piece') is classified negatively by the absence of a generic meaning and lack of the formal markers of Nouns or Verbs. The full definition of the 'Particle' usually includes the qualification 'which occurs for some [syntactic] meaning' (*jā'a li-ma'nā*) to separate it from other linguistic units also called *ḥarf,* which would nowadays be labeled 'phoneme', 'grapheme', 'lexeme', and other nonsyntactic units.

(3) The functional categories reveal the ultimate preoccupation of Arabic linguistics with speech as behavior. There are more than 70 terms for the various ways in which words can be used, ranging from the decisions of the speaker at the highest level (e.g. *ibtidā'* 'starting an utterance' [specifically a statement of the form 'x is y', where Arabic has no verb for 'is'], *istifhām*

'asking a question') to the specific events occurring within the utterance as a result of grammatical rules (e.g. *naṣb* 'making dependent' [scil. 'accusative' or 'subjunctive'], *taṯniya* 'making dual'). The precise boundary between the speaker's autonomy and submission to grammatical rules is never very clear, but this is intentional, as both the speaker and speech elements can be 'operators' (*'āmil*, see above), i.e. affect the form of words. It is noteworthy that the 70 or more names of speech acts are all expressed in Arabic by Verbal Nouns ('inflecting', 'making diminutive', 'describing [adjectivally]', 'eliding', 'doubling [a sound]' etc.), so that Arabic grammar, like an Islamic legal treatise, is nothing but a list of actions and conditions for their permissibility.

The many functions enable words of the same form class to be subcategorized: thus, Adjectives and Adverbs are simply Nouns used in a certain 'way' (*naḥw*) or 'place' (*mawḍi'*), and conversely Demonstrative and Relative Pronouns, which do not have the form of Nouns, nevertheless have the 'status' (*manzila*) of Nouns and are thus labeled 'noun of pointing' (*ism 'išāra*) and 'noun of the attached [clause]' (*ism al-mawṣūl*), respectively. The western Prepositions, Interjections, Conjunctions, and Articles are all considered Particles (*ḥarf*), and are categorized by the function they express; hence, a Vocative Particle is called a *ḥarf nidā*', 'a bit for calling', a Negative Particle is a *ḥarf nafy*, 'a bit for negating', a [coordinating] Conjunction is a *ḥarf 'aṭf*, 'a bit for joining', and so on. In this way, every speech element, from sound to sentence, is identified and its behavior is accounted for.

The above sketch is based on the description of Arabic by the first grammarian, Sībawayhi (who died in late eighth century AD), which remains unsurpassed for its completeness and scientific adequacy. Later grammar added almost nothing to the data but made substantial changes to the theory as Islamic culture absorbed the methods of Greek science, leading in one direction to a marked pedagogical and scholastic character and in another to increased abstraction and growing interest in underlying features. What might be called the metaphysical aspects of linguistics, such as the origins of language itself, the nature of meaning, and the question of grammatical causality, were enthusiastically and revealingly explored but, as might be expected, they had to be dealt with in a manner that did not conflict with Islamic doctrines.

References

Auroux, Sylvain, E.F.K[onrad] Koerner, Hans-Josef Niederehe, and Kees Versteegh (eds.) 2000. Handbook of the language sciences, Vol. 1. Berlin, New York: Walter de Gruyter.

Bakalla, Muhammad H. 1983. Bibliography of Arabic linguistics, 2nd edition. London: Mansell.

Bohas, George, Jean-Patrick Guillaume, and Djamel E. Kouloughli. 1990. The Arabic linguistic tradition. London and New York: Routledge.

Carter, Michael G. (ed.) 1981. Arab linguistics, an introductory classical text with translation and notes. Amsterdam/Philadelphia: John Benjamins.

Carter, Michael G. 1991. 'Lexicography' and 'grammar'. The Cambridge history of Arabic literature, Vol. 1, Religion, learning and science in the Abbasid Period, ed. by Michael J. Young, et al. Cambridge: Cambridge Univerity Press.

Carter, Michael G. 2004. Sībawayhi. Oxford: Oxford University Press/1. B. Tauris.

Gully, Adrian. 1994. Grammar and semantics in medieval Arabic. Richmond: Curzon.

Sezgin, Fuat M., Geschichte des arabischen Schrifttums, Vol. 8, 'Lexikographie', Vol. 9, 'Grammatik'. Leiden: E.J. Brill, 1981, 1984.

Versteegh, Kees. 1995. The explanation of linguistic causes. Al-Zaǧǧāǧī's theory of grammar. Introduction, translation, commentary. Amsterdam/Philadelphia: John Benjamins.

Versteegh, Kees. 1997. The Arabic linguistic tradition. Landmarks in linguistic thought III. London and New York: Routledge.

Michael G. Carter

See also **Arabic**

Aramaic

Aramaic is a northwest Semitic language, related to Hebrew and more distantly to the now extinct Akkadian, which it replaced around the third century BCE as the lingua franca of the middle East, and to Arabic, South Arabian, and Ethiopic.

The nomenclature of Aramaic is quite confusing. Aramaic is sometimes called Chaldee, Chaldaic, or Assyrian. The latter appellation is also used for one of the dialects of Akkadian, and it should not be confused with that different language. In Israel, Aramaic is

sometimes called Kurdish, because Aramaic speakers living there hail from Kurdistan, but the term Kurdish is properly applied to a quite different western Iranian language. An important eastern dialect of Aramaic is usually called Syriac. Syriac has three native scripts that have a one-to-one relationship with the standard Hebrew alphabet, which was itself originally used for Aramaic and pushed out the native Hebrew alphabet during the Babylonian exile. Jewish dialects of Aramaic are normally written in Hebrew characters. Aramaic is indigenous to Syria and is known from the tenth century BCE, but it has been in use at different times from Egypt to India. It has remained a living language in a few areas up to now. Thus, in 1930, in the town of Zakho in Iraqi Kurdistan, there were approximately 1,500 Aramaic-speaking Jews. All of these people and their descendants have moved to Israel, where their language lives on tenuously. There are other small pockets, particularly in Maalula, near Damascus, Syria, where the language is spoken mostly by Greek Catholics. Other related dialects are spoken by Christians in the Tur Abdin area of southern Turkey. Aramaic must nevertheless be considered an endangered language, constantly pressed by Arabic, Hebrew, Turkish, and other languages, and it seems unlikely that there are any monolingual speakers now. In the year 2000 election in the United States, the city of Chicago published instructions for voter registration in various languages, including Aramaic, and offered a telephone hotline where Aramaic speakers could call for information in their own language. There is a section in that city where the shops have Aramaic signs and advertisements. Aramaic is also quite widely used as a liturgical language by ancient Christian churches, including the Chaldean Catholic church, and that of four groups of Malabar Christians in India, known as the Christians of St. Thomas. The latter group has been replacing its Aramaic liturgy with its vernacular Malayalam language. The Samaritans, a Jewish sect that has existed from biblical time until the present day, seem to have spoken Aramaic until about the tenth century CE and possessed texts in the language, but now they speak Arabic. A number of important prayers in the Jewish liturgy are still recited in Aramaic, particularly the *Kaddish* (Doxology) and the *Kol nidre* (annulment of vows).

Aramaic is of crucial importance for religious studies. The biblical books of Daniel and Ezra have substantial portions written in Aramaic, and there is one verse in Jeremiah [10:11], as well as two words in the book of Genesis [31:47.] The last item displays the disparity in the lexicon between Hebrew and Aramaic: Jacob calls a monument Gal-Ed ['heap of witness': the story explains the name of the area east of Jordan called Gilead], whereas Laban calls it Jegar-Sahadutha, which is a literal translation into Aramaic. Additionally, two of the four so-called Ancient Versions of scripture are in Aramaic: the Targum ('translation'), which is a Jewish translation of the Hebrew Scriptures into Aramaic, and the Peshitta ('simple version'), a Christian translation of the Bible into the dialect of Aramaic called Syriac. These versions give important evidence of variant readings and show how obscure passages were understood in olden times. The Talmud, originally oral in nature, is written predominantly in Aramaic, with an admixture of Hebrew. The city of Edessa, now called Sanliurfa, in southeastern Turkey, was once a center of Christian religious schools that produced a vast literature in the Syriac dialect, including philosophy, theology, and poetry, much of which exists in unpublished manuscripts in various scholarly libraries. The scriptures of the Mandaean religion, a dualistic faith that still exists in Iraq and Iran, are written in Mandaic, which is an Aramaic dialect, and adds yet another name to the language.

Various Aramaic inscriptions exist on stone, clay, and other writing materials from the period before 700 BCE over a wide area from Asia Minor to northern Arabia. The importance and international character of the language may be gauged from the fact that when the Akkadian-speaking Assyrians besieged Jerusalem at this time, the officials on the wall begged the emissary to say what he had to say in Aramaic, because they understood it, but did not wish the common people to understand. The emissary refused and spoke his threats as to the horrors of the siege in Hebrew, to discourage the defenders (II Kings 18:26). A particularly important discovery from a later period is a collection of documents, all in Aramaic and written on papyrus, found at the ancient city of Yeb in Egypt, whose ruins are on an island called Elephantine opposite the modern city of Aswan. It seems that there was a Jewish temple in Yeb from the seventh century BCE, and there was a community of Aramaic-speaking Jewish mercenaries who left this substantial evidence of their existence. The documents contain official and private letters, legal documents, and some literary works. Of particular historical importance is a copy of a decree by the Achaemenid king Darius II Nothus, who reigned in the second half of the fifth century BCE, instructing the Jews of Elephantine to observe the Passover. This lent credence to the Aramaic decrees found in the book of Ezra, which some had thought to be forgeries. Another celebrated document is a request for permission to rebuild the magnificent Jewish temple of Yeb, which had been destroyed around 410 BCE by the priests as well as various legal documents, that shed a flood of light on the history of the time.

Aramaic has a large number of dialects, but all have the features characteristic of Semitic languages. Words

of the core vocabulary contain a root, usually of two, three, or four consonants, to which brief prefixes and suffixes are added, and they also undergo internal vowel change; many of these roots have equivalents in other Semitic languages. Nouns have two genders, masculine and feminine, and two numbers, singular and plural. The verb is inflected with suffixes in the perfect tense and with prefixes and suffixes in the imperfect tense, like other Semitic languages. Aramaic numerals above two share the common Semitic peculiarity of using the feminine form with masculine nouns, and vice versa.

Literary western Syriac is a clearly defined standard language. The 22 letters of its alphabet represent the consonants of the language. The vowel system distinguishes /a/,/a:/,/e/, /i/ and /u/; /u/ and /o/ were not distinctive in this dialect. For an illustration of word formation, consider the word [ri:ʃ] 'head.' Pronominal suffixes may be added to nouns, of which this is a selection: [ri:ʃe:h] 'his head', [ri:ʃa:n] 'our head', [ri:ʃkun] 'your (plural) head'. Various nouns have irregularities; thus, the plural of [bra:] 'the son' is [bnaya:], the plural of [ħa:ta:] 'the sister' is [ʔaħwa:ta:], and the plural of [ʔatta:] 'woman, wife' is [neʃe:].

The fact that the participle in Aramaic is not used to construct complex sentences with frequent subordinate clauses, as occurs, say, in Germanic, Slavic, or Dravidian, means that Aramaic syntax is very simple compared with that of those languages. The verb usually comes first, followed by the subject and object, and sentences are linked together by conjunctions. The result is a language that is mostly regular and straightforward, with a minimum of verb tenses, no case endings on nouns, and a native script that serves it well.

References

Brown, Leslie. 1982. The Indian Christians of St Thomas. Cambridge: Cambridge University Press.

Driver, G.R. 1957. Aramaic documents of the fifth century B.C. Oxford: Clarendon Press.

Heinrichs, Wolfhart. 1990. Studies in Neo-Aramaic. Atlanta: Scholars Press.

Jastrow, Marcus. A dictionary of the Targumim. New York: Putnam, 1886–1903 frequently reprinted.

Johns, Alger F. 1966. A short grammar of Biblical Aramaic. Berrien Springs, ml: Andrews University Press.

Lindenberger, James M. 1994. Ancient Aramaic and Hebrew letters. Atlanta: Scholars Press.

Maclean, Arthur J. 1971. Grammar of the dialects of vernacular Syriac. Amsterdam: Philo Press.

Macuch, R. 1965. Handbook of classical and modern Mandaic. Berlin: De Gruyter.

Marcus, David. 1981. A manual of Babylonian Jewish Aramaic. Lanham: University Press of America.

McNamara, Martin (project director). 1986. The Aramaic Bible.Wilmington, NC: Glazier.

Robinson, Theodore H.1962. Paradigms and exercises in Syriac grammar. Oxford: Clarendon Press.

Stevenson, William B. 1962. Grammar of Palestinian Jewish Aramaic. Oxford: Clarendon Press.

Wright, William. 1966. A short history of Syriac literature. Amsterdam: Philo Press.

ALAN D. CORRÉ

Arawak

The Arawak language family is the South American language family with the largest number of languages. Geographically, it spans four countries of Central America—Belize, Honduras, Guatemala, Nicaragua—and eight of South America—Bolivia, Guyana, French Guiana, Suriname, Venezuela, Colombia, Peru, and Brazil (and also formerly Argentina and Paraguay).

There are about 40 living Arawak languages. The first Native American peoples encountered by Columbus—in the Bahamas, Hispaniola, and Puerto Rico—were Arawak-speaking Tainos. Their language became extinct within a hundred years of the Columbian invasion. Spanish—and many other languages—inherited a number of loanwords from Arawak languages. These include widely used words such as *hammock, tobacco, potato, guava,* and many other names for flora and fauna.

The creation of a 'mixed' language of Arawak/Carib origin in the Lesser Antilles is one of the most interesting pieces of evidence for the history of languages in pre-Conquest times. Speakers of Iñeri, a dialect of the Arawak language now (misleadingly) called Island Carib, were conquered by Carib speakers. They developed a 'mixed' Carib/Arawak pidgin that survived until the seventeenth century (Hoff 1994). 'Men's speech' and 'women's speech' were distinguished in the following way. Women used words of Arawak origin, while men used words of Carib origin and grammatical elements mostly of Arawak origin. The pidgin coexisted with Carib used by men and Iñeri used by women

and children; it belonged to both parties and served as a bridge between them. This diglossia gradually died out with the spread of Island Carib to both men and women. As a result, Island Carib, an Arawak language, had a strong influence of Carib with respect to both vocabulary and, possibly, grammar.

The languages in areas settled by the European invaders soon became extinct. Those on the north coast of South America perished first, before 1700. When the search for gold and rubber extended up the Amazon and its tributary the Rio Negro, additional languages succumbed, a process continuing from the eighteenth century to the present day. Sometimes, the Indians retaliated, attacking settlements and missions; but the invaders always returned. Indian rebellions often provoked forced migrations that sometimes ended up in the creation of a new dialect or even language. For instance, in 1797 the British authorities removed the rebellious inhabitants of St. Vincent (an island in the Lesser Antilles) to Belize on the mainland. These were racially a mixture of black slaves and Indians, who spoke Island Carib. This resulted in the creation of a new dialect of Island Carib—known as Central American Island Carib, Kariff, Black Carib, or Garifuna—which by the twentieth century had developed into a separate language, now one of the two Arawak languages with the largest number of speakers.

The overwhelming majority of Arawak languages are now endangered. Even in the few communities with over 1,000 speakers, a national language (Portuguese or Spanish) or a local lingua franca (Lingua Geral Amazônica, Quechua, or Tucano) is gradually gaining ground among younger people. The few healthy Arawak languages include Guajiro in Venezuela and Colombia (probably up to 300,000 speakers) and Garifuna in Central America (with up to 190,000 speakers). The Campa languages (total estimate 40–50,000) form one of the largest groups of indigenous population in Peru.

Most of the materials on Arawak languages collected during the second half of the twentieth century are by missionary linguists. Their quality and quantity vary. A full description is available for only three or four languages.

The fact that Arawak languages were related was first recognized by Father Gilij as early as in 1783. The recognition of the family was based on a comparison of pronominal prefixes in Maipure, a now extinct language from the Orinoco Valley, and Moxo, from Bolivia. Gilij named the family Maipure. Later, it was renamed Arawak by Daniel Brinton after one of the most important languages of the family, Arawak (or Lokono), spoken in the Guianas. This name gained wide acceptance during the following decades. The majority of native South American scholars use the

TABLE 1 Pronominal Prefixes and Suffixes in Proto-Arawak

Person	Prefixes		Suffixes	
	Singular	Plural	singular	plural
1	*nu-* or *ta-*	*wa-*	*-na, -te*	*-wa*
2	*(p)i-*	*(h)i-*	*-pi*	*-hi*
3nf	*ɾi-, i-*	*na-*	*-ɾi, -i*	*-na*
3f	*thu-, ru-*	*na-*	*-thu, -ru, -u*	*-na*
'Impersonal'	*pa-*	—	—	—

name Arawak (Aruák) to refer to the group of unquestionably related languages easily recognizable by shared pronominal prefixes such as *nu-* or *ta-* '1sg', *pi-* 'you', and prefixes *ka-* meaning 'have' and *ma-* meaning 'not'. A number of scholars, mainly North Americans, prefer to use the term Arawak(-an) to refer to a much more doubtful higher-level grouping and reserve the term Maipuran (or Maipurean) for the group of undoubtedly related languages, which was claimed to be one branch of 'Arawakan'. Here, I follow South American practice and use the name Arawak for the family of definitely related languages.

The limits of the family were established by the early twentieth century. Problems still exist concerning internal relationships within the family and possible relationships with other groups. Internal classification and subgrouping of Arawak languages remains a matter of debate; further detailed work is needed on both the descriptive and comparative fronts.

The putative studies of 'Arawakan' by Ester Matteson, G. Kingsley Noble, and others are deeply flawed. Unfortunately, these have been adopted as the standard reference for the classification of Arawak languages, especially among some anthropologists, archeologists, and geneticists, influencing ideas on a putative homeland and migration routes for proto-Arawakan. The classification found in Campbell (1997) contains a number of factual mistakes and omissions and has to be treated with extreme caution.

Little is known about a homeland for the Arawak family. The linguistic argument in favor of an Arawak homeland located between the Rio Negro and the Orinoco rivers—or on the Upper Amazon—is based on the fact that there is a higher concentration of structurally divergent languages found in this region. This area has also been suggested as one of the places where agriculture developed. This is highly suggestive and corroborated by a few mythical traditions of northern origin found among Arawak-speaking peoples south of the Amazon. The origin myths of the Tariana, in northwest Amazonia, suggest that they could have come from the north coast of South America.

Arawak languages are complicated in many ways. Words can be differentiated by stress in some

languages, such as Baure and Waurá (south of Amazonas), and Tariana, Achagua, and Warekena (north of Amazonas). At least two use tones, i.e. pitch differences, to distinguish words—Terêna in the south, and Resígaro spoken in the far northeast of Peru.

Each Arawak language has a few prefixes and numerous suffixes. Prefixes are typically monosyllabic, while suffixes can consist of one or more syllables. Roots usually contain two syllables. Prefixes are rather uniform across the family, while suffixes are not. What is an independent word in one language can be a grammatical suffix in another language. An Apurinã noun *maka* means 'clothing'—this is where the English word *hammock* comes from. In Tariana, *-maka* is a classifier for clothing, as in *pa:-maka* 'one piece of clothing' (literally 'one-classifier: clothing').

Most grammatical categories in Arawak languages are expressed on the verb. Cases for marking subjects and objects are atypical. Tariana, spoken in northwest Brazil, has developed cases for such basic grammatical relations to match the pattern in nearby East Tucano languages; cases are also reported for Apurinã.

Arawak languages spoken south of the Amazon ('South Arawak') have a more complex verb structure than those north of the Amazon ('North Arawak'). South Arawak languages such as Amuesha or Campa have up to 30 suffix positions. North Arawak languages such as Tariana or Palikur do not have more than a dozen. Suffixes express meanings that are realized by independent words in familiar Indo-European languages, e.g. 'be about to do something', 'want to do something', 'do late at night', 'do early in the morning', 'do all along the way', 'in vain', 'each other'. A typical example from Amuesha, spoken in Peru, shows the complexity of meaning that verbal suffixes can express *ø-omaz-amy-eʔ-amp^y-es-y-e.s-n-e.n-a* (3person.singular-go.downriver-distributive-epenthetic-dative-epenthetic-plural-epenthetic-late-progressive-reflexive) 'They are going downriver by canoe in the late afternoon stopping often along the way'.

Arawak languages spoken north of the Amazon have evidentials, i.e. forms with which speakers specify how they acquired the information. Baniwa and Piapoco have just a 'reported' evidential, which is used for second-hand information. Tariana has four evidentials: visual (if one saw the event happen), nonvisual (if one heard or smelt it), inferred (if one can judge what happened by some indirect evidence), and reported (if one acquired the information from another person).

Verbs are divided into active verbs ('hit' or 'jump') and stative verbs ('be cold'). All Arawak languages share pronominal affixes and personal pronouns. Pronominal suffixes (e.g. *-ni* 'he') are used for subjects of stative verbs and for direct objects. Subjects of active verbs are expressed by prefixes (e.g. *nu-* 'I').

For instance, in Baniwa, one says *nu-kapa* 'I see' and *nu-watsa* 'I jump', but *nu-kapa-ni* 'I see him' and *hape-ni* 'he is cold' (*nu-* refers to 'I' and *-ni* to 'him'). And 'my hand' is *nu-kapi*.

Some languages have lost the pronominal suffixes; these include Yawalapiti (Xingú park, Brazil) and Chamicuro (Peru) to the south of the Amazon, and Bare, Resígaro, Maipure, and Tariana, to the north. The form of the first person pronoun 'I' is *ta-* in the Caribbean (Lokono, Guajiro, Añun, Taino) and *nu-* in other languages. This is the basis for classification of Arawak languages into *Nu*-Arawak and *Ta*-Arawak.

Most Arawak languages distinguish two genders— masculine and feminine, e.g. Palikur *amepi-yo* 'thief (woman)' (literally 'thief-MASCULINE'), *amepi-ye* 'thief (man)' (literally 'thief-FEMININE'), Tariana *nu-phe-ri* 'my elder brother' (literally 'I-sibling-MASCULINE'), and *nu-phe-ru* 'my elder sister' (literally 'I-sibling-FEMININE'). Genders are not distinguished in the plural. Some languages also have complicated systems of classifiers—these characterize the noun in terms of its shape, size, or function. Tariana and Baniwa have over 40 classifiers, e.g. Tariana *pa:-da* (one-classifier: round) 'one round thing', *hanu-da* (big-classifier: round) 'big round thing', i.e. the suffix *-da* is a classifier for 'round things' and can be attached to numerals ('one'), adjectives ('big'), etc.

All Arawak languages distinguish singular and plural. Plural is only obligatory with human nouns. Plural markers are **-na/-ni* 'animate/human plural', **-pe* 'inanimate/animate non-human plural'.

Throughout the Arawak language family, nouns divide into those that must have a possessor (inalienably possessed) and those that do not require a possessor as a must (alienably possessed). Inalienably possessed nouns include body parts, kinship terms, and a few others, e.g. 'house' or 'name'. Inalienably possessed nouns have an 'unpossessed' form marked with a suffix, e.g. Pareci *no-tiho* 'my face' (literally 'I-face'), *tiho-ti* '(someone's) face' (literally 'face-UNPOSSESSED'). Alienably possessed nouns take a suffix when possessed, e.g. Baniwa *nu-činu-ni* 'my dog' (literally 'I-dog-POSSESSED').

The overwhelming majority of Arawak languages have a negative prefix *ma-* and its positive counterpart *ka-*, e.g. Piro *ka-yhi* 'having teeth' (literally 'POSITIVE-tooth'), *ma-yhi* 'toothless' ('NEGATIVE-tooth'); Bare *ka-witi-w* 'a woman with good eyes' ('POSITIVE-eye-FEMININE'), *ma-witi-w* 'a woman with bad eyes, a blind woman' ('NEGATIVE-eye-FEMININE'). Most languages have just the numbers 'one' (*pa-;* also meaning 'someone, another') and 'two' (*(a)pi* or *yama*).

The common Arawak vocabulary consists mostly of nouns, with quite a few body parts, fauna, flora, and artifacts. Only a few verbs are retained in most Arawak

languages, e.g. *ikau* 'arrive', **pɨ(da)* 'sweep', **po* or **da* 'give', **(i)ya* 'cry', **kama* 'be sick, die', **itha* 'drink'.

References

Aikhenvald, Alexandra Y. 1995. Person-marking and discourse in North-Arawak languages. Studia Linguistica 49. 152–95.

———. 1996. Areal diffusion in North-West Amazonia: the case of Tariana. Anthropological Linguistics 38. 73–116.

———. 1998. Warekena. Handbook of Amazonian languages, Vol. 4, ed. by D.C. Derbyshire and G.K. Pullum, 215–439. Berlin: Mouton de Gruyter.

———. 1999a. The Arawak language family. In Dixon and Aikhenvald, pp. 65–105.

———. 1999b. Areal diffusion and language contact in the Içana-Vaupés basin, North West Amazonia. In Dixon and Aikhenvald, pp. 385–415.

———. 2001. Areal diffusion, genetic inheritance and problems of subgrouping: a North Arawak case study. Areal diffusion and genetic inheritance: problems in comparative linguistics, ed. by Alexandra Y. Aikhenvald and R.M.W. Dixon, 167–94. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. 2002. Language contact in Amazonia. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. 2003. A grammar of Tariana, from northwest Amazonia. Cambridge: Cambridge University Press.

———, and Diana Green. 1998. Palikur and the typology of classifiers. Anthropological Linguistics 40. 429–80.

Campbell, Lyle. 1997. American Indian languages. New York: Oxford University Press.

Dixon, R.M.W., and Alexandra Y. Aikhenvald. 1999. Introduction. In Dixon and Aikhenvald, pp. 1–21.

Dixon, R.M.W., and Alexandra Y. Aikhenvald (eds.) 1999. The Amazonian languages. Cambridge: Cambridge University Press.

Facundes, Sidney da Silva. 2000. The language of the Apurinã people of Brazil. Ph.D. Thesis. SUNY at Buffalo.

Hoff, Berend. 1994. Island Carib, an Arawakan language which incorporated a lexical register of Cariban origin, used to address men. Mixed languages: 15 case studies in language intertwining, ed. by P. Bakker and M. Mous, 161–8. Amsterdam: IFOTT.

Payne, David L. 1991. A classification of Maipuran (Arawakan) languages based on shared lexical retentions. Handbook of Amazonian languages, Vol. 3, ed. by D.C. Derbyshire and G.K. Pullum, 355–499. Berlin: Mouton de Gruyter.

Taylor, Douglas M. 1977. Languages of the West Indies. Baltimore, MD: Johns Hopkins University Press.

Tovar, Antonio, and Consuelo Larrucea De Tovar. 1984. Catálogo de las lenguas de América del Sur. Madrid: Editorial Gredos.

Wise, Mary Ruth. 1986. Grammatical characteristics of Preandine Arawakan languages of Peru. Handbook of Amazonian languages, Vol. 1, ed. by D.C. Derbyshire and G.K. Pullum, 567–642. Berlin: Mouton de Gruyter.

ALEXANDRA AIKHENVALD

See also **Carib and Cariban Languages**

Archeology and Language

The history of related languages is at the same time a history of societies to which those languages belonged. When we reconstruct prehistoric relationships among a group of languages, we simultaneously postulate the historical existence of the societies that spoke those languages. Since the very beginnings of linguistics as a science in the nineteenth century, the problem of the archeological correlates of language change and dispersal has thus been crucial to the joint efforts of linguists and archeologists.

It was within the field of Indo-European linguistics that archeological evidence has been more and more widely resorted to determine the location of the Indo-European homeland, the area in which a population spoke Proto-Indo-European, a broadly defined group of dialects that later developed into the various Indo-European languages. The homeland problem can be stated as follows: there are several reasons for supposing that the historical distribution of Indo-European languages is not the original one and that the Indo-European language family, at some time in its prehistoric existence, has occupied a territory far more confined than that of its earliest historically attested branches; Indo-European languages are thus supposed to have spread from somewhere to their historical location in prehistory, invading or occupying areas where languages not belonging to the Indo-European family were spoken; some of these relic non-Indo-European languages are attested throughout Eurasia (Basque, Tartessian, Iberian, Etruscan, Dravidian languages), and are surrounded by Indo-European languages. To some extent, the archeological record can suggest the sociohistorical conditions under which the Indo-Europeanization of Eurasia may have occurred. From the great number of archeological solutions to the homeland problem, one can select three scenarios that enjoy wide currency:

(1) *The continuity hypothesis*. Archeologists have found that there is a major cultural border between the steppe cultures and those of

Northern/Central Europe, which was not seriously transgressed at any period from the Neolithic period onward. Thus, a Proto-Indo-European linguistic continuum should already have existed during the Mesolithic (8500–5000 BC) between the Baltic and the Black/Caspian seas. The Neolithic cultures that emerged throughout this region are then to be considered as ancestral to their respective regional Indo-European languages.

(2) *The Anatolian hypothesis*. According to this hypothesis, the dispersal of Indo-Europeans is connected with the spread of agriculture from Anatolia into Europe and Asia: Indo-European farming colonists absorbed non-Indo-European cultures expanding in a wave of advance.

(3) *The Kurgan model*. According to this model, the Indo-European homeland was located among pastoralist tribes of south Ukraine and south Russia that expanded both to the east and west (the *kurgan* or tumulus being a typical marker of this expansion). This model implies a process of progressive acculturation of non-Indo-European peoples across Europe and Asia.

Archeological models of dispersal have also been applied to other areas. In Australia, patterns of linguistic diffusion show considerable similarities to the patterns of diffusion of nonlinguistic traits. For instance, both the use of dug-out canoes and pearl shell ornaments appear to have permeated down from the coastal regions of northern Australia as far as South Australia, a pattern that closely resembles the south-

ward diffusion of some linguistic traits. In Africa, significant steps have been made toward uncovering the early historical developments among the speakers of each of the major language families of the continent: Nilo-Saharan, Niger-Congo, Afro-Asiatic, and Khoisan. The Proto-Nilo-Saharan speech territory most probably lay in the southern fringe of the central Sahara desert. The early periods of Khoisan history have been played out in eastern Africa: Khoisan speakers were makers of the Eastern African Microlithic Tradition; as early as the sixth millennium BC, Khoisan speakers began to spread south across southern Africa. Proto-Afroasiatics are to be identified with the early wild-grass-collecting cultures of northeastern Africa. The overall history of the dispersal of Niger-Congo languages is exceedingly complex, but an initial stage (between 8000 and 6000 BC) can be clearly identified in which Niger-Congo peoples, starting as far east as the Nuba Mountains of Sudan, expanded south across the woodland savannas of West Africa.

References

Dixon, R.M.W. 2002. Australian languages. Their nature and development. Cambridge: Cambridge University Press.

Ehret, Christopher. 2000. Language and history. African languages. An introduction, ed. by Bernd Heine and Derek Nurse. Cambridge: Cambridge University Press.

Gimbutas, Marija. 1991. The civilization of the goddess. San Francisco: Harper.

Renfrew, Colin. 1987. Archeology and language. London: Jonathan Cape.

ANDREA SANSÒ

See also **Indo-European 1: Overview**

Armenian

Armenian is a cover term for a number of distinct languages, the best known of which are the literary forms Standard Eastern and Western Armenian, Middle (/Medieval/Cilician) Armenian, and Classical Armenian. The label 'Armenian' also encompasses the closely related languages Zok, formerly spoken by the Armenian inhabitants of southeastern Nakhichevan; Kistinək, spoken by the Armenian inhabitants of Musaler, Turkey; Kesbənuək, spoken by the Armenian inhabitants of Kesab, Syria; Homshetsma, spoken by the Hemshinli of northeast Turkey and the Hamshen Armenians of the Black Sea coastal regions of

Abkhazia and Russia; and dozens of other mutually unintelligible variants of Armenian originally spoken in Turkey, Armenia, Azerbaijan, Iran, Georgia, Abkhazia, Russia, and Israel. Lomavren, the language of the Bosha gypsies of Turkey and Armenia, draws its phonology, morphology, and syntax from the Erzerum dialect of Armenian but its lexicon is mostly of Indic origin; for this reason, it is not clear whether or not the language should be classified as a member of the Armenian family. All but Homshetsma employ the Armenian alphabet created by Mesrob at the beginning of the fifth century.

Armenian is an Indo-European language, generally believed to be most closely related to the Greek and Indo-Iranian subgroups. (For instance, all three share a particle *me: used in expressing prohibitions (Greek me:, Sanskrit ma:, Armenian mi) and the imperfect third-person singular particle *e- (as in Greek e-pher-e, Sanskrit a-bhar-a-t, Armenian e-ber '(s)he/it carried').) Because of the large number of loan words from various Middle Iranian languages, especially Parthian, Armenian was thought to be an Iranian dialect until Heinrich Hübschmann demonstrated in 1875 that it was a distinct branch of the Indo-European family. Scholars disagree on how the Armenians came to historical Armenia, the eastern half of present-day Turkey centered around Lake Van and Mount Ararat. Some believe they came southward from the Russian steppe, others believe they and the Hittites came eastward from Greece, and others suggest they moved only a short distance from an original Indo-European homeland in the Transcaucasus. It is most likely that this settlement occurred in the second millennium BC. The earliest mentions of the Armenians occur in the inscriptions of the Achaemenid Persian king Darius (sixth century BC) and the Greek historian Herodotus (fifth century BC).

The earliest written records of the Armenian language date from the fifth century AD, shortly after the conversion of the Armenians to Christianity in the fourth century led to the creation of an Armenian alphabet by Mesrob in around 401 and a systematic program of translating the books of the Bible. The language of the earliest translations was Classical Armenian (also called grabar, 'written (language)'), which continued as the preferred literary form of Armenian until the nineteenth century, when it was supplanted by the two modern literary dialects.

In linguistic terms, Armenian is notable for its significant divergences from Proto-Indo-European (the reconstructed ancestral language of the Indo-European family), particularly in terms of pronunciation and vocabulary. Some of the more striking phonological changes are the development of a rich set of affricates (ts, dz, etc.), the loss of final syllable rimes (e.g. Proto-Indo-European *worgʲom 'work' > Classical Armenian gorts), the change of initial *dw to erk- (e.g. Proto-Indo-European *dwo: '2' > Classical Armenian erku), and the change of original *w to g. Most striking in the vocabulary of Armenian is the rarity of words inherited from Indo-European, and the overwhelming predominance of words of unknown origin. Not surprisingly, native Indo-European words survive primarily in the core vocabulary: mayr 'mother' < *ma:ter, hayr 'father' < *pater, kʰoyr 'sister' < *swesor, kov 'cow' < *gʷows, tun 'house' < *domos, em 'I am' < *esmi. The remainder of the lexicon is drawn primarily from Parthian, and to a lesser extent from Greek and Syriac; several hundred and perhaps as many as several thousand words are of unknown origin, most likely having come from Urartian, Hurrian, and other now-extinct autochthonous languages. Armenian also incorporated large numbers of Arabic words following the expansion of the Arabs in the Middle East in the seventh century, and thousands of Turkish words following the arrival of the Seljuks and other Turkic tribes in Anatolia beginning in the eleventh century.

Although there are dozens of mutually unintelligible varieties of Armenian, all share certain features. Proto-Armenian had four verbal conjugations, characterized by theme vowels -e-,-i-,-ɑ-, and -u- (beɾ-e-m 'I carry', χaws-i-m 'I speak', χnd-a-m 'I rejoice', and zgen-u-m 'I wear'); most modern dialects (including the Western and Eastern literary languages) have completely or partially lost the -u- conjugation, and standard Eastern Armenian has merged the -i- conjugation into the -e- conjugation. There were originally three morphologically distinct sets of personal endings for verbs—present, imperfect, and aorist—which were used in combination with additional tense and aspect markers to form the various tenses and moods. The system of nominal morphology in Proto- and Classical Armenian was rich, preserving the Indo-European nominative, accusative, genitive, dative, instrumental, ablative, and locative cases in both singular and plural (but the Indo-European dual was lost); there were at least eight different declensions, distinguished primarily by different theme vowels. This system was significantly reduced by the medieval period; Middle Armenian and the modern varieties now use the singular endings for the plural as well, and have only one productive declension, formed from parts of the original -i- and -o- declensions. With the exception of pronouns, the inventory of cases has significantly reduced as well: the accusative has merged with the nominative, and the genitive with the dative. Proto-Armenian had several participial forms, but only two of these survive into the modern period: the original past participle -eal is now -el in the Eastern dialects, and the original present participle -oʁ is now used as a present participle and for relativizing subjects of subordinate clauses. The Western dialects have replaced -eal with -ats (> -adz in most varieties) for past participles; all modern dialects also use the -ats participle to relativize nonsubjects of subordinate clauses.

Most of the changes between Classical and Modern Armenian first appeared in the medieval period in Middle Armenian documents, associated with the Armenian kingdom of Cilicia, which flourished from the eleventh to fifteenth centuries AD in what is now south-central Turkey. Middle Armenian is generally

Western in character, although it shares many features with Eastern dialects as well. It inverts the pronunciation of the Classical Armenian plain voiced and voiceless stops (e.g. berem 'I carry' > perem, pat 'wall' > bad), a feature that is preserved in the modern Cilician dialects of Zeytun and Hadjin but differs from the Western and Eastern literary varieties (Eastern preserves the Classical system (bɛrɛm); Western devoices and aspirates the original voiced consonants (pʰɛrɛm)). The Cilician kingdom was in close contact with several Crusader kingdoms, as a result of which it borrowed a significant number of words from Crusader French, most famously what comes out as the standard Western form for 'mister,' *baron*.

In the nineteenth century, Armenian nationalists became interested in developing a literary form of the modern language. This was brought about by excising most Turkish forms from the regional dialects and replacing them with new borrowings from the classical language. The intellectual center around which the new Western literary language was organized was Constantinople, although many features of the standard dialect (including the pronunciation of the consonants) do not come from the Armenian dialect originally spoken there. The same holds for Eastern Armenian with respect to Erevan. The relationship between the two modern literary dialects is somewhat complicated; there are many grammatical differences (e.g. Western gə sirɛm vs. Eastern sirum ɛm 'I love', Western bidi sirɛm vs. Eastern kəsirɛm 'I will love') and lexical differences (e.g. Western dʒɛrmag vs. Eastern spitak 'white'; Western hɔs vs. Eastern ɛstɛʁ 'here', Western bɛdkʰaran vs. Eastern zukʰaran 'bathroom', Western havgitʰ vs. Eastern dzu 'egg'), and most Western speakers have difficulty in understanding Eastern; however Eastern speakers are relatively comfortable with the Western dialect. This asymmetry in mutual intelligibility most likely results from the fact that large numbers of speakers of Western dialects fled to Eastern Armenia following the Russo-Turkish war in 1828 and the Turkish Genocide in 1915–1920, whereas before the fall of the Soviet Union in 1991 most Western Armenians had little or no exposure to Eastern Armenian. The fact that there is some mutual intelligibility in both directions can also be linked to the fact that the literary dialects tend to borrow the same forms from Classical Armenian, and (at least in recent decades) employ the same newly coined words.

The destruction of the Armenian homeland and more than a million Armenians by the Ottoman government in 1915–1920 rendered most nonstandard varieties of modern Armenian moribund; with few exceptions, the Armenians in the diaspora (primarily Lebanon, France, and the United States) speak only Standard Western Armenian. There were approximately 6.8 million speakers of Armenian in 1996, but all varieties of the language except for Standard Eastern Armenian are in immediate danger of extinction as very few diaspora Armenians under the age of 30 speak the language fluently.

Whereas Classical Armenian was relatively Indo-European in its syntactic and morphological structure, all varieties of Modern Armenian are typologically much closer to Turkish and the Balkan languages. Compare, for instance, the formation of relative clauses, exemplified by 'I saw the bird that was singing in the tree': Classical tesi əz-tʰərtʃʰun-ən or ɛrgɛr i vɛraj tsar-oj-n (I.saw specific-bird-definite that was.singing in on tree-genitive-definite), Western dzar-i-n vəra jɛrkʰɔʁ tʰərtʃʰun-ə dɛsa (tree-genitive-definite on singing bird-definite I.saw). Western Armenian has been additionally influenced by Turkish and Greek (cf. əsdɛpʁin 'carrot', istakʰɔz 'lobster', bantʰɔg 'hotel'), whereas Eastern Armenian has been heavily influenced by Russian (e.g. the standard form for 'potatoes' is kʰartʰɔfli, and the word for 'gay' is galubɔj, from the Russian word originally meaning 'sky blue'; the native word for 'blue,' kapujt, cannot be used in this sense).

The Armenian Alphabet, with IPA Equivalents for Eastern Pronunciation

ա	ɑ	ծ	ts	ջ	dʒ
բ	b	կ	k	ռ	r
գ	g	հ	h	ս	s
դ	d	ձ	dz	վ	v
ե	ɛ	ղ	ʁ	տ	t
զ	z	ճ	tʃ	ր	r
է	ɛ	մ	m	ց	tsʰ
ը	ə	յ	j	ու	u
թ	tʰ	ն	n	փ	pʰ
ժ	ʒ	շ	ʃ	ք	kʰ
ի	i	ո	ɔ	o	ɔ
լ	l	չ	tʃʰ	ֆ	f
խ	χ	պ	p		

The Lord's Prayer in Different Varieties of Armenian, Rendered in the IPA

Classical Armenian *(Ējmiacin ms. 229, 989 A.D.)*

Հայր մեր որ յերկինս. սուրբ եղիցի անուն քո. եկեսցէ արքայութիւն քո. եղիցին կամք քո որպէս յերկինս եւ յերկրի. զՀաց մեր հանապազորդ տուր մեզ այսաւր. եւ թող մեզ զպարտուիս մեր. որպէս եւ մեք թողումք մերոց պարտապանաց. եւ մի տանիր զմեզ ի փորձութիւն. այլ՝ փրկեա զմեզ ի չարէ. զի քո է արքայութիւն եւ զաւրութիւն եւ փառք յաւիտեանս ամէն:

hajr mer or jɛrkinəs. surb ɛʁitsʰi anun kʰo. ekɛtsʰɛ arkʰajutʰiwn kʰo. ɛʁitsʰin kamkʰ kʰo orpɛs jɛrkinəs ew jɛrkri. əzhatsʰ mer hanapazord tur mez ajsawr. ew tʰoʁ mez əzpartis mer. orpɛs ew mekʰ tʰoʁumkʰ merotsʰ

partapanats^h. ew mi tanir əzmez i p^hordzut^hiwn. ajʁ^j p^hərkea əzmez i tʃ^harɛ. zi k^ho ɛ ark^hajut^hiwn ew zawrut^hiwn ew p^hark^h jawiteanəs amɛn.

Standard Eastern Armenian

Հայր մեր, վոր յերկնքում ես. սուրբ թող լինի քո անունը. քո թագավորությունը թող գա. քո կամքը թող լինի յերկրի վրա, ինչպես վոր յերկնքում է. մեր Հանապազօրյա Հացը տուր մեզ այսօր. յեվ թող մեզ մեր պարտքերը, ինչպես յեվ մենք ենք թողնում մեր պարտականներին. յեվ մի տար մեզ քործւթյան, այլ փրկիր մեզ չարից. վորովՀետեվ քոնն է թագա–վորությունը յեվ զորությունը յեվ փառքը Հավիտյանս. ամէն:

hajɾ mɛr, vɔr jɛrkənk^hum-ɛs. surp^h t^hɔʁ lini k^hə anunə. k^hə t^hagavɔrut^hjunə t^hɔʁ ga. k^hə kamk^hə t^hɔʁ lini jɛrkri vəra, intʃ^hpɛs vɔr jɛrkənk^hum ɛ. mɛr hanapazɔrja hatsʰə tur mɛz ajsɔr. jɛv t^hɔʁ mɛz mɛr partk^hɛrə, intʃ^hpɛs jɛv mɛnk^h ɛnk^h t^hɔʁnum mɛr partakannɛrin. jɛv mi tar mɛz p^hɔrdzut^hjan, ajl p^hərkir mɛz tʃ^haritsʰ. vɔrɔvhɛtɛv k^hɔn: ɛ t^hagavɔrut^hjunə jɛv zɔrut^hjunə jɛv p^hark^hə havitjanəs. amɛn.

Standard Western Armenian

Ով Հայր մեր որ երկինքն ես, քու անունդ սուրբ ըլլա. քու թագաւորութիւնդ գա. քու կամքդ ըլլա ինչպես երկինքը՝ նոյնպես երկրի վրայ. մեր ամէն օրուան Հացը այսօր ալ մեզի տուր, մեզի ներէ մեր պարտքերը ինչպես մենք ալ կը ներենք մեր պարտականներուն. ու մեզ փորձւթեան մի տանիր, Հապա չարէն մեզ ազատէ. քանզի քուկդ է թագաւորութիւնը եւ զորութիւնը ու փառքը յաւիտեանս: Ամէն:

ɔv hajɾ mɛr vɔr jɛrgink^hn-ɛs, k^hu anunət^h surp^h əl:a. k^hu t^hak^havɔrut^hjunət^h k^ha. k^hu gamk^hət^h əl:a intʃ^hbɛs jɛrgink^hə, nujnbɛs jɛrgri vəra. mɛr amɛn ɔrvan hatsʰə ajsɔr al mɛzi dur, mɛzi nɛrɛ mɛr bardk^hɛrə intʃ^hbɛs mɛnk^h al gə nɛrɛnk^h mɛr bardagan:ɛrun. u mɛz p^hɔrtsʰut^hjan mi danir, haba tʃ^harɛn mɛz azadɛ. k^hanzi k^hugət^h ɛ t^hak^havɔrut^hjunə jɛv zɔrut^hjunə u p^hark^hə havidjanəs. amɛn.

Zeytun Dialect (Cilicia, South Central Turkey)

ɔv mɛj bobə ɔj ijgink^hn-is, k^hu anunət sujp t^hɔʁna. k^hu t^hɛk^hɛvyyt^hynət t^huʁ kɔ. k^hu gɔmk^hət t^huʁ la, intʃ^hbɛs ijgink^hə, indɛn ɛl ijgɛjin vijɔ. mij amɛn œjvɛn hɔtsʰə ɛsœj miz tuj. jɛv miz nɛjɛ mij bɔjdk^hə, tʃ^hotsʰ vɔr mink^h ɛl gə nɛjink^h mij bɔjdk^hi dɛjɛrun. jɛv miz p^hɔjtsut^han mi danəj, habə tʃ^hɔjen miz azadɛ. tʃ^hunk^hi k^hin: ɛ t^hɛk^hɛvyjyt^hynə jɛv zɔjut^hynə u p^hɔrk^hə. havidjanəs havidɛnitsʰ. amɛn.

Kesab (Syria)

œv miɛr bybə, surp ɛʁni k^hɛ ænun, k^hɛ t^hɛk^hɛvyrut^hynə t^hɔʁ kɔ, k^hɛ iradɛt^hət ənːɔ, tʃ^hytsʰər k^hi irgænk^hə t^hərzɛn ɛl ikɛdinə, miɛr amɛn ɛvyr hɔɔtsʰə dur miɛz ɛs ɛvyr ɛl, miɛr bɔrdk^hə miɛzi baʁətʃ^hlamuʃ əɾɔ tʃ^hytsʰər k^hi miɛnk^h ginonk^h miɛrontsʰə, vɛ zəzmiɛz p^hɔrtsyt^hjan mi danɔ, habə χaləsɔ i tʃ^harien, tʃ^hynk^hi k^hɛ ɛ t^hɛk^hɛvyrut^hynə, tʃ^hɛrɛfə, k^huvɛt^hə, havidieinəs havidɔnitsʰ amɔn.

References

Adjarian, Hrachea. 1909. Classification des dialectes arméniens. Paris: Champion.

Bardakjian, Kevork, and Bert Vaux. 1998. Eastern Armenian: a textbook. Ann Arbor: Caravan.

Clackson, James. 1994. The linguistic relationship between Armenian and Greek. Oxford: Blackwell.

Dumézil, Georges. 1964. Notes sur le parler d'un Arménien musulman de Hemşin. Brussels: Palais des Académies.

Finck, Franz Nikolaus. 1907. Die Sprache der armenischen Zigeuner. St. Petersburg: Imperial Science Academy.

Karst, Josef. 1901. Historische Grammatik des Kilikisch-Armenischen. Strassburg: Trübner.

Meillet, Antoine. 1936. Esquisse d'une grammaire comparée de l'arménien classique, 2nd edition. Vienna: Imprimerie des Pères mékhitharistes.

Thomson, Robert. 1975. An introduction to classical Armenian. Delmar, NY: Caravan.

Vaux, Bert. 1998. The phonology of Armenian. Oxford: Oxford University Press.

BERT VAUX

See also **Indo-European 1: Overview**

Artificial Intelligence

Artificial Intelligence (AI) is the subfield of computer science that aims to enhance computers with capabilities traditionally attributed to human intelligence, such as perception, reasoning, understanding, and learning. However, AI researchers usually give more elaborate definitions for AI, focusing either on the problems that

AI aims to solve or on the methods used for approaching them. Thus, AI can be alternatively defined as the effort to make computers capable of solving problems that currently only humans can solve, or as a set of methods for confronting problems for which exact solution algorithms are not yet known. The latter definitions also explain why topics such as symbolic calculus and neural networks are not considered pure AI topics anymore and why some not-well-defined problems, such as natural language understanding, are traditionally positioned in the domain of AI.

AI has been strongly influenced by the works of Norbert Wiener, Claude Shannon, John von Neumann, Alan Turing, and Alonzo Church. However, the term 'artificial intelligence' is attributed to John McCarthy, who introduced it on August 31, 1955, in a document titled *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.* The proposal involved a two-month study 'on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it'. It was signed by John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon. The conference that was held at Dartmouth College (in Hanover, New Hampshire) the next summer is considered a milestone in the history of AI because it brought together all the researchers interested in the field and established the birth of AI. After the Dartmouth conference, Minsky and McCarthy founded the AI Laboratory at the Massachusetts Institute of Technology, and two other participants—Herbert Simon and Allen Newell— founded the AI Laboratory at Carnegie Mellon University. McCarthy, Minsky, Simon, and Newell are considered the founders of modern AI.

The feasibility of the ultimate AI goal (i.e. the development of intelligent machines) has been questioned repeatedly. A theorem known as the Church–Turing thesis has provided strong supporting evidence, assuming that intelligent behavior is based on some kind of computation. Turing has also described a process to test whether the intelligence of an AI system is comparable to that of a human being. This process is known as the Turing test: a human observer is allowed to interact via a teletype with another human and a machine that pretends to be a human. The observer does not know which is which and tries to figure it out by conversing with both via the teletype. Turing argued that if the machine could successfully fool a knowledgeable observer, then it should certainly be considered intelligent. However, many philosophers, such as John Searle and Hubert Dreyfus, have argued not only against the correctness of the Turing test but also against the feasibility of ever building a nonbiological intelligent machine. This debate led to the formation of two philosophical directions: strong AI and weak AI. Supporters of strong AI believe that computers, even with current technology, are capable of supporting intelligence, given appropriate programming. Supporters of weak AI believe that although programs able to imitate intelligent behavior can be made, it is impossible to make a machine that understands or develops consciousness in the way that humans do.

The distinction between weak and strong AI is closely related to another distinction: that between symbolic and subsymbolic AI. The key assumption of symbolic AI is that knowledge is represented by structures of semantically meaningful symbols and that a set of such symbols is therefore required to support the reasoning process of an intelligent system. A program for proving logical theorems is an example of a symbolic AI system. However, the key assumption of subsymbolic AI is that intelligent behavior can be attained without the need for semantically meaningful symbols. A neural network is an example of a subsymbolic AI system.

A generic AI system can be described in terms of four main components: perception, reasoning, knowledge, and learning. Figure 1 shows their interconnection. Perception is the task of processing the input to extract structured information needed for the particular task. The reasoning component is responsible for the manipulation of this information, according to the system knowledge contained in the knowledge component. The output of reasoning may be the outcome of the AI system itself, or it may be fed to a learning mechanism
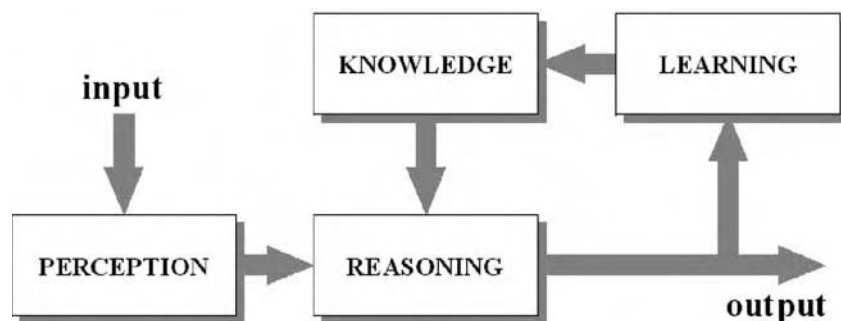
Figure 1. Structure of a Generic Artificial Intelligence System.

whose function is to update the existing knowledge. The structure of Figure 1 should be considered as a descriptive model, not an implementational one, because some systems may lack some components, and others, especially the subsymbolic ones, may have some of the components merged indistinguishably.

However, the majority of the topics addressed by AI concern one or more of the above four components. These topics are discriminated as AI branches, according to the types of problems, the techniques used, and the classes of applications involved. There are many AI branches: search methods, symbolic reasoning, production systems, heuristic techniques, knowledge representation, logic programming, statistical reasoning, fuzzy logic, game playing, planning, vision, natural language processing, machine learning, neural networks, genetic algorithms, common sense reasoning, expert systems, speech understanding, theorem proving, machine translation, robotics, perception, automatic programming, constraint satisfaction, ontology, and intelligent agents, among others. Their outcomes are very diverse. Some of them, such as expert systems and game playing, have produced impressive results, whereas others, such as machine translation and speech understanding, are still far from their goal.

There are several scientific organizations devoted to AI research. Among the best known are the American Association for Artificial Intelligence (AAAI), the European Coordinating Committee for Artificial Intelligence (ECCAI), the Society for Artificial Intelligence and Simulation of Behavior (AISB), and the Special Interest Group on Artificial Intelligence (SIGART) of the Association for Computing Machinery (ACM).

References

Barr, Avron, and Edward A. Feigenbaum (eds.) 1981. The handbook of artificial intelligence. Los Altos, CA: William Kaufmann Inc.

Haugeland, John (ed.) 1997. Mind design II. Cambridge, MA: MIT Press.

Jackson, Philip C. Jr, 1985. Introduction to artificial intelligence, 2nd edition. New York: Dover Publications.

Konar, Amit. 2000. Artificial intelligence and soft computing: behavioral and cognitive modeling of the human brain, Boca Raton, FL: CRC Press.

Nilsson, Nils J. 1998. Artificial intelligence: a new synthesis. Los Altos, CA: Morgan Kaufmann.

Pollock, John L. 1989. How to build a person: a prolegomenon. Cambridge, MA: MIT Press.

Rich, Elaine, and Kevin Knight. 1991. Artificial intelligence, 2nd edition. New York: McGraw-Hill.

Russel, Stuart, and Peter Norvig. 1995. Artificial intelligence: a modern approach. Eaglewood Cliffs, NJ: Prentice–Hall.

Turing, Alan M. 1950. Computing machinery and intelligence. Mind 59. 433–60.

KYRIAKOS N. SGARBAS

See also **Turing, Alan**

Artificial Languages

The term 'artificial language' can refer to several different kinds of symbolic systems, including programming languages, symbol systems used in various branches of science, 'controlled' languages for unambiguous technical communication, intermediate languages in machine translation, highly regularized ethnic or literary languages, fictional languages created for literary purposes, and planned languages for intercultural communication. In linguistics, the term is most often used to designate this last category, but metaphors and concepts from these neighboring fields are a frequent source of both insight and confusion.

The idea of improving existing languages, by making them more regular, less ambiguous, or more expressive arose in medieval Europe. By the mid-seventeenth century, a number of leading intellectuals, including Jan Amos Komensky (Comenius), René Descartes, Gottfried Leibnitz, and John Wilkins, were investigating the possibility of constructing a linguistic system on the basis of scientific taxonomies and logicomathematical operations. This work inspired many so-called philosophic or a priori language projects throughout the next two centuries, both written (*pasigraphies*) and spoken (*pasilalies*). Despite their influence on the development of mathematics, formal logic, and general linguistic theory, none of these systems acquired a significant body of users.

In the nineteenth century, the rise of historical and evolutionary views of language contributed to a waning of interest in philosophical language projects. At the same time, improvements in transport and communications turned attention toward the practical benefits of a common international language, while the Romantic equation of language with culture and identity led

many to associate the idea of a common language with ideals of peace and brotherhood among nations. Today, all three of these orientations—the philosophical-scientific, the instrumentalist, and the value oriented—can be found among users, researchers, and theorists of international planned languages.

The systematic study of such languages, a field now referred to as *interlinguistics*, was inaugurated in 1903 with the publication of Couturat and Leau's *Histoire de la Langue Universelle* (*History of the Universal Language*), supplemented four years later by *Les Nouvelles Langues Universelles* (*The New Universal Languages*). The authors analyzed a total of 75 language projects and project outlines, most of them originating in the previous half-century. No fewer than 912 projects are listed in the most complete catalogue to date, Aleksandr Duličenko's *Meždunarodnye Vspomogatel'nye Jazyki* (1990: *International Auxiliary Languages*). However, the vast majority of these have seen no practical use, and linguistic research on them is necessarily theoretical. Empirical research requires a community of language users, and the number of projects that has achieved this status is quite small.

The most widely used classificatory scheme in interlinguistics places language projects on a continuum between 'naturalism' (modification of an existing language that leaves most of its features intact) and 'schematism' (implying a high degree of rationalization on a priori lines). All of the more naturalistic projects that have acquired a community of users are Latinate. They include Latino sine flexione (1903), a regularized form of Latin invented by the Italian mathematician Giuseppe Peano; Occidental (1922; later renamed Interlingue), the project of Edgar de Wahl, an Estonian mathematician who had been active in earlier projects; Novial (1928), the creation of Danish linguist Otto Jespersen; and Interlingua (1951), a project published under the auspices of the International Auxiliary Language Association, but primarily designed by its director Alexander Gode. Like the earlier naturalistic projects that it replaced, Interlingua was intended primarily for communication among educated Europeans; Gode's idea was to distill a 'Standard Average European' (a concept inspired by Benjamin Whorf) from lexical and grammatical elements common to English, French, Italian, and Spanish/Portuguese. Interlingua acquired perhaps a few thousand users in the 1950s and 1960s, primarily in the United States and Europe, and retains a small following today.

The remaining language projects to have acquired a body of users display both naturalistic and schematic traits. They are also quite diverse. The earliest of them, Volapük (1879), was developed by Johann Martin Schleyer, a Catholic priest in Baden, southern Germany. The vocabulary of Volapük consisted of arbitrary modifications of Western European words, making new coinages controversial. After its extraordinarily rapid spread in the 1880s, including a certain amount of oral use, Volapük underwent an even more rapid decline in the 1890s in response to internal conflicts over the control and development of the language. It has not been in active use since the 1920s.

Volapük's successor, Esperanto (1887), was published by a Jewish physician in Warsaw, Ludovik Zamenhof. Initially slower to spread than its predecessor, Esperanto experienced explosive growth between 1900 and 1912, the period in which its oral tradition became firmly established. Today, it remains by far the mostly widely used international planned language, and thus it is the most accessible to linguistic study. For much the same reason, it is also the language in which the greatest share—approximately 50%—of the interlinguistic literature has been written. Research on Esperanto is reviewed in greater detail below.

In 1907, Ido, a project derived from Esperanto, was published and quickly attracted some prominent adherents from the Esperanto movement; rivalry between the two languages continued until after World War II. It is now generally accepted that Ido was the creation of Louis Couturat, a prominent French logical philosopher and the principal author of *Histoire de la Langue Universelle* (mentioned previously). Some of Couturat's modifications tended toward greater Latinization. Others introduced greater schematism. Like the Volapük community before it, the Ido movement soon became embroiled in linguistic disputes that alienated many potential users. Today, only a very small community of speakers remains.

The newest language projects to have acquired a small number of users are Glosa, a derivative of Lancelot Hogben's Interglossa (1943), and Lojban, a derivative of James Cooke Brown's Loglan (1955). Glosa combines Greek and Latin words with particles to denote tense, number, and so forth; phrase and sentence construction resemble those of English. Lojban's syntax is based on symbolic logic; it was originally proposed, in the tradition of the philosophic language projects, as a means of exploring the relationship between language and thought, and attempts to use it for oral communication are fairly recent. Neither project has acquired a speech community.

All but the last two projects are discussed in greater detail in Detlev Blanke's seminal work *Internationale Plansprachen* (1985; *International Planned Languages*), the standard reference work on the subject. This work introduced an alternative classificatory scheme, listing 18 stages through which a language project might be expected to pass on its way to full socialization. On this scale, Ido and Interlingua were identified as 'semilanguages', having developed beyond

the project stage to acquire an original and translated literature, oral use at international meetings, and other markers of communicative use. Esperanto, however, was shown to be a qualitatively different phenomenon, being used on a scale and in a range of situations as extensive as, although different from, many ethnic languages.

Zamenhof's original blueprint for Esperanto consisted of 16 short rules on a variety of topics, a glossary of 940 words, some samples of poetry and prose, and a lengthy essay on the key characteristics of his proposal. Like most planned languages, Esperanto uses the Latin alphabet, and pronunciation follows orthography; in words with several syllables, intonational stress falls on the next to last syllable. Chief among its distinguishing features is its system of word-class endings: *-o* marks nouns, *-a* adjectives, and *-e* adverbs; *-i/-is/-as/-os/-us/-u* form verbs in the infinitive, past, present, and future indicative, conditional, and imperative modes, respectively. Apart from certain closed class words (the definite article *la*, prepositions, conjunctions, comparatives, correlatives, and a small number of adverbs ending in *-aŭ*), all lexical items must be combined with one of these endings to function in a sentence. Word roots are borrowed intact or with inflectional endings removed, primarily from Romance (75%), but also from Germanic (20%), Slavic, and other sources. Direct loans from non-European languages remain relatively rare.

A second key characteristic of Esperanto is the invariance and high combinability of all of its elements. Zamenhof introduced more than 20 affixes to express ideas related to the root word, such as 'place of', 'container for', 'opposite of', and so on; these have since been supplemented by a handful of others. Word roots may also be freely combined. Many common compounds have acquired conventional meanings over time, so that *lern-ej-o,* for instance (literally 'learning-site'), usually indicates a school of some kind. Early compound formation was influenced by Germanic and Slavic models (e.g. *el-don-i,* to publish, literally 'to give out'), but modern usage favors greater transparency and what might be called 'indigenous' models, e.g. the direct transformation of adjectives into verbs (*bon-a,* 'good', *bon-i,* 'to be good, to manifest goodness') or the use of affixes and prepositions as independent roots (*ad-e,* 'continually, repeatedly', *per-i,* 'to mediate, broker', *et-a,* 'small').

This introduces a third key characteristic of Esperanto: its evolution through communicative use. Throughout 25 years as the central figure of the movement, Zamenhof consistently emphasized an inductive approach to learning and using the language, with new speakers developing their own tacit linguistic models on the basis of a growing body of literary and scientific texts. This approach is reflected in the exercises of the *Fundamento* (1904; *Foundation of Esperanto*), which Zamenhof persuaded the first international Esperanto congress (1905, Boulogne-sur-Mer) to adopt as an unalterable paradigm of acceptable usage. It also accounts for the effort he invested in translations of such works as Gogol's *The Tax Inspector,* Shakespeare's *Hamlet,* the Old Testament (from Hebrew), and Andersen's *Fables*; these influential texts did much to develop Esperanto literary style and semantics. After World War I, literary and scientific writing in Esperanto flourished and diversified in both book and periodical form; major libraries currently hold 10,000 to 20,000 volumes. Development of the oral language has also been marked, particularly since the 1970s, with the expansion of international telephony and travel. Today, although influential dictionaries, grammars, and textbooks exist, none can be taken as a definitive or complete description of Esperanto as it is actually used.

Valuable research on Esperanto has been carried out both by professional linguists and philologists and by brilliant amateurs such as the Hungarian polymath Kálmán Kalocsay (*Plena Gramatiko de Esperanto,* 1932; *Complete Esperanto Grammar*). Only a small sample of this work is available in English. Particular interest has been focused on the rules governing word and sentence formation; the etymology and semantics of the core vocabulary; the interplay of prescriptive principles (such as the 16 rules), linguistic ideologies, and tacit linguistic models; the relationship between language, culture, and identity; the development of Esperanto literature; and the use of the language in various scientific and technical fields. In the 1980s, a Dutch project examined Esperanto's suitability as an intermediate language for machine translation. The Modern Languages Association of America lists 200 to 300 publications in its annual bibliography.

Several million people have studied Esperanto, on the basis of textbook sales and other data; however, the population of those with an active command of the language is much more difficult to characterize. Taking into account the reported membership of the Universal Esperanto Association (the most representative international organization of Esperanto speakers), annual sales of books and periodicals, and economic barriers to participation, a reasonable current estimate would be something more than 50,000 active speakers, along with a substantially larger number of passively competent, lapsed, or occasional users. Some three quarters of active speakers are located in Europe (including Eastern Europe and Russia), with established communities of users in most East Asian and Latin American countries, in West Africa, and in regionally isolated countries such as Iran and Israel. There are active, self-organizing networks of both young speakers (under 30 years) and Esperanto-speaking families, including a

few hundred first-language speakers. Active specialist organizations exist in many areas, including predominantly oral domains such as popular music and radio broadcasting. Increasingly, the World Wide Web constitutes a valuable tool for research into the language and its domains of use.

References

Blanke, Detlev. 1985. Internationale Plansprachen: Eine Einführung (International planned languages: an introduction). Berlin: Akademie-Verlag.

Couturat, Louis, and Léopold Leau. 1979. Histoire de la langue universelle (1903), Les nouvelles langues universelles (1907). Hildesheim and New York: Olms.

Duličenko Aleksandr. 1990. Meždunarodnye vspomogatel'nye jazyki (International auxiliary languages). Tallinn: Valgus.

Eco, Umberto. 1993. La ricerca della lingua perfetta. Roma: Laterza; as The search for the perfect language, translated by James Fentress. Oxford: Blackwell, 1995.

Forster, Peter G. 1982. The Esperanto movement. The Hague: Mouton.

Gledhill, Christopher. 1998. The grammar of Esperanto: a corpus-based description. München/Newcastle: Lincom Europa, 2nd edition, 2000.

Janton, Pierre. Espéranto. 1973. Paris: Presses Universitaires de France; as Esperanto: language, literature, and community, ed. by Humphrey Tonkin, translated by Humphrey Tonkin, Jane Edwards, and Karen Johnson-Weiner. Albany: State University of New York Press, 1993.

Jordan, David. 1992. Being colloquial in Esperanto: a reference guide for Americans. Lanham: University Press of America, 2nd edition, El Cerrito: Esperanto League for North America, 1999.

Large, Andrew. 1985. The artificial language movement, Oxford: Blackwell.

Schubert, Klaus (ed.) 1989. Interlinguistics: aspects of the science of planned languages. Berlin: Mouton de Gruyter.

Tonkin, Humphrey (ed.) 1997. Esperanto, interlinguistics, and planned language. Lanham: University Press of America.

MARK FETTES

Aspect

'Aspect' in its proper sense—also called 'grammatical aspect'—is one of the grammatical categories of the verb besides tense, mood, and others. The term 'aspect' is also used as a cover term for grammatical aspect and two other closely related phenomena: 'aktionsart' as a lexical-semantic category of verbs and 'predicational aspect' as a semantic property of verb phrases. Aspect in its broader sense has to do with the internal structure of the event the verb refers to, with the perspective the speaker adopts toward the course and structure of the event, or with both of these.

Because the expression of grammatical aspect is often closely intertwined with the expression of tense, in traditional grammars, aspectual distinctions are often confusingly dealt with under the heading of 'tense'. Among the categorical distinctions in tense-aspect systems that are of an aspectual nature are those between *imparfait* and *passé composé/passé simple* in French, progressive and nonprogressive in English, or imperfect and aorist in Greek. Formally, aspect can be expressed by very different means, e.g. by prefixes and suffixes (Russian), reduplication (Chamorro), inflectional and periphrastic forms within a tense-aspect system (French, English), verb stem formation (Greek), or combinations thereof (Bulgarian).

The main opposition within the grammatical category of aspect is the one between perfective and imperfective aspect. Informally, it is often said that in using the perfective aspect, the speaker looks onto the event from outside as an undivided whole, often focusing on its completion. From the imperfective perspective, the speaker views the event from inside as an ongoing process, a state, or a habitual action without directing attention to the completion of the event. From a textual point of view, verbs in the imperfective aspect contribute to the description of the background of a scene, whereas verbs in the perfective aspect are used to describe the events that take place on this background in their temporal succession. Thus, the Russian equivalent for English *to write* has an imperfective form, *pisat',* and a perfective one, *napisat'*. The imperfective one is used in contexts in which English would use expressions such as *she was writing* or *she used to write*. The perfective form occurs in contexts in which the completion or the result of writing is emphasized (*he wrote/has written the letter*) and in which it occurs in a sequence of events (*he sat down, wrote the letter, and put it in an envelope*). Aspectual semantics is a very intricate matter, and the meaning of the perfective–imperfective opposition involves cross-linguistic differences. Furthermore, some languages do not show a fully developed perfective–imperfective distinction but express aspectual meanings that can roughly be classified as meaning components of the imperfective aspect (progressive vs. nonprogressive aspect, habitual vs. nonhabitual aspect) or the perfective aspect (completive vs. noncompletive aspect).

In contrast to grammatical aspect, the term 'aktionsart'—which translates as 'kind of action' from German—is considered to be a lexical-semantic category. Approaches to aktionsart differ as to whether they consider all verbs to show a certain aktionsart or whether they restrict the term to phenomena in which aktionsart meanings are expressed by certain prefixes or suffixes. Aktionsart distinctions are mainly used to classify verbs according to their reference to parts of events, to events of a certain length, and to repetitions of events (affixes and their translations are underlined): German *loslaufen* 'to start running', Russian *posidet'* 'sit for a while', Russian *kriknut'* 'to scream once', German *austrinken* 'to drink up', Russian *pisyvat'* 'to write repeatedly'. Because both grammatical aspect and aktionsart are mainly expressed by means of word formation in Slavic languages—which have played an important role in research on aspect—the proper distinction between the two categories has always been a subject of debate. The diversity of the phenomena covered by the term 'aktionsart' is reflected in the diverse semantic approaches to aktionsart. It can be observed, though, that formal semantic approaches to aktionsart phenomena often involve notions like 'part of an event' (an event referred to by German *austrinken* 'to drink up' as part of an event referred to by *trinken* 'to drink'), 'recurrence of an event' (*rewrite, rebuild*), and 'event presupposition' (*he ate up* presupposes *he ate*; *she rebuilt the house* presupposes *somebody built the house before*, etc.)

Sometimes, the term 'aktionsart' is also applied to or—better—confused with the phenomenon of 'predicational aspect' (also called 'lexical aspect'). The most influential description of predicational aspect was given by Zeno Vendler (1957), who distinguished four 'time schemata:' (1) statives like *know something, hate somebody*; (2) achievements (roughly, expressions referring to punctual events) like *reach the summit, win a race, recognize somebody*; (3) activities (roughly, expressions referring to events as ongoing or non-culminating) like *push a cart, run,* and *swim*; and (4) accomplishments (roughly, expressions referring to

events as culminating or delimited) like *draw a circle, run a mile, write a letter*. Despite the popularity of these so-called Vendler classes, they are problematic in many respects. In particular, the four classes are distinguished by semantic properties belonging to very different domains (length of event, stativity, completion), and it is often unclear if the domain of the classification is verbs, verb phrases, or sentences.

Formal semantics has been especially interested in the difference between activities (e.g. *she ate, she ate apples, she watched the apples*) and accomplishments (e.g. *she ate the apples*). Expressions describing activities can be modified with phrases like *for ten minutes*, but not phrases like *in ten minutes*. Because *watch the apples* is an activity, *She watched the apples for ten minutes* is acceptable, but *She watched the apples in ten minutes* is not. With accomplishments, such as *eat the apples*, it is the other way around: *She ate the apples in ten minutes* is acceptable, whereas *She ate the apples for ten minutes* only seems to be so if the context forces *eat the apples* to be reinterpreted as an activity.

References

Binnick, Robert I. 1991. Time and the verb. A guide to tense and aspect. New York, Oxford: Oxford University Press.

Chung, Sandra, and Alan Timberlake. 1985. Tense, aspect, and mood. Language typology and syntactic description. Vol. III: Grammatical categories and the lexicon, ed. by Timothy Shopen. Cambridge: Cambridge University Press.

Comrie, Bernard. 1976. Aspect. An introduction to the study of verbal aspect and related problems. Cambridge: Cambridge University Press.

Dahl, Östen. 1985. Tense and aspect systems. Oxford: Basil Blackwell.

Vendler, Zeno. 1957. Verbs and times. The Philosophical Review LXVI. 143–60.

Verkuyl, Henk J. 1993. A theory of aspectuality. The interaction between temporal and atemporal structure. Cambridge: Cambridge University Press.

STEFAN ENGELBERG

See also **Inflection and Derivation; Tense and Aspect Marking; Time and Tense**

Assimilation and Coarticulation

Assimilation refers to a change in sound due to the influence of a neighboring sound (aspiration, glottalization, nasal release, etc.). For instance, the /s/ in 'husband' becomes voiced as a result of the voiced phonemes surrounding it. The influence can be either in

anticipation of the next sound (= anticipatory or right-to-left assimilation), or it can be carry-over (left-to-right) assimilation, in which an ongoing feature is continued to the next sound. Assimilation is similar to coarticulation, in that both processes deal with the mod-

ification of sound due to contextual variability. However, the standard generative phonology (Chomsky and Halle, the Sound Pattern of English, 1968) makes a clear-cut distinction between assimilation and coarticulation. Assimilatory processes are part of a grammar and are language specific (although they occur in many languages), and they can be accounted for by phonological rules. Coarticulation results from the inertia of the speech mechanism and is, hence, more universal. Both the distribution (universality vs. language specificity) and the quality of the contextual change (mere articulatory adaptation vs. intentional phonetic modifications) distinguish the two processes. However, there is no simple way of distinguishing between those assimilation effects that are due to the inherent properties or limitations of speech production and those that are not, unless the latter are very obviously language-specific. Assimilation often appears to be motivated by ease of articulation, but what seems easy and natural in one language often turns out to be less so in another.

Assimilation effects are not fully understood: it is not clear to what extent we can explain them by assuming that high-level commands associated with specific segments are confounded by biomechanical 'sloppiness' or to what extent high-level commands are deliberately planned to optimize transitions between targets and to yield the best possible vocal tract performance in running speech.

Assimilation is often mentioned in connection with historical changes. Many of the sound changes in languages can be described as assimilatory. For example, English words such as mission, special, or passion were once pronounced with a medial [sj] but in modern English have [ʃ]: by a process of assimilation, the [s] has been retracted in anticipation of the following [j] or [I]. However, it is important to distinguish between historical processes (which are over and done with) and processes that are still current or operative in the modern language. They are part of the speaker's organization of pronunciation and are relevant to our description of the language (e.g. prefix un-, which is pronounced as [n] in nonvelar contexts (untidy, unsettled) but which are pronounced as [ñ] before a velar (unkind, ungainly).

Place of Articulation

This occurs in English, for example, by 'ratbag' or 'oatmeal' pronounced with [p] instead of [t] in rapid or informal speech, by assimilation of the alveolar stop to a following bilabial.

Manner of Articulation

Here, 'Indian' is pronounced as 'Injun', where the stop [d] and approximant [j] merge to form an affricate (same for soldier).

Voicing

This occurs, for instance, when 'v' in 'have to' is pronounced with an [f] instead of [v] due to assimilation of the voiced fricative to the following voiceless consonant.

Not all speech sounds are equally affected by contextual variability. Bilabial, [b] and [p], tend to be very stable and show no effect of assimilation irrespective of the context. Velars such as [k] and [g] are also stable, although the place of articulation is more fronted in the context of front vowels: the closure of the [k] in 'kick' is more palatal, while that of the [k] in 'cook' is more velar. Of the different plosives, assimilation occurs most in alveolars, i.e. in [t] and [d]. The closure of the [t] or [d] will not be alveolar, but will occur at the place of articulation of the following segment. For instance, the closure of a [t] or [d] before a dental will also sound dental (indicated by a diacritic), e.g. 'ho[t̪ θ]ing 'hot thing', while that before a velar will sound velar, i.e. 'ho[k k]ake', 'hot cake' or 'ba[g g]irl', 'bad girl.' Preceding a bilabial, the [t] or [d] will also be bilabial, i.e. 'hot pie' becomes 'ho[p p]ie' and 'bad boy' becomes 'ba[b b]oy'.

Manner of Assimilation

Alveolar stops are not only influenced by surrounding consonants but also by surrounding vowels. Here, the assimilation involves manner rather than place of articulation. In several North American accents, the [t] and [d] are neutralized, i.e. their characteristic physical properties are replaced by the voiced alveolar flap (transcribed as [ɾ]). This sound contains voicing and a flap (= a very brief contact between the tongue tip and the alveolar ridge). Using this flap, the words 'Adam' and 'atom' may sound identical, both words having the flap for the intervocalic 't' and 'd'. This flapping occurs whenever what would be [t] or [d] in other accents occurs between two vowels, not only within words (see previous example) but also across word boundaries, as in 'ge[ɾ] away' (get away). However, flapping does not occur when the stop precedes a stressed syllable, as in 'attract', where the second syllable carries the stress.

Affricates

There is little assimilation of the affricates in English, although there may be some variation among speakers.

Labiodental fricatives [f] , [v] do not show a great deal of assimilation, although [v] may become a voiceless word, finally preceding a voiceless obstruent (ha[f] to, o[f] course.

In English, in fast speech, dental fricatives [ð] often assimilates entirely to a preceding alveolar sound 'I[n n]epub 'in the pub'.

The alveolar [s] and [z] are often assimilated to the following palatal glide [j] or palatoalveolar fricative as in 'mi[š j]ou' ('miss you'), 'it wa[ʒ j]ellow' ('it was yellow').

Assimilation is quite a general phenomenon in connected speech, but it spreads wider in fast speech than in slow speech. It involves the change of some segment under the influence of another one, and the change makes the two speech sounds more similar. The phrase 'ten books', for example, is pronounced as [tem buks], where /n/ assimilates to the adjacent /b/ by adopting its bilabiality feature. Most forms of assimilation are to be located at the segment spellout level.

Assimilation should be distinguished from coarticulation. Adjacent speech sounds interact because of the physiology and the mechanics of articulation. These interactions become more intense at higher speech rates. They depend on the time allotted to the articulation of syllables (coarticulation).

There is a wide range of assimilation processes that take place across a syllable or word boundary. In Swedish, a syllable final /r/ is integrated with any (phonemically adjacent) apical dental consonant(s) making the latter a retroflex alveolar, and this effect spreads over the word boundary.

Quite often, context-dependent changes involving the same articulatory structures have different acoustic and perceptual manifestations in different languages, so that it is possible to distinguish what can be considered universal phonetic behavior from language particular rules. A classical example is the difference between vowel harmony, an assimilatory process present in a limited number of languages (Hungarian) and the process of vowel to vowel coarticulation, attested in many languages and probably present in all. In other cases, cross-language differences are not easily interpretable and inferences about the nature of the underlying processes can be made only by manipulating some of the speech parameters, for example segmental durations. In a study of vowel nasalization in Spanish and American English, Solé and Ohala (1991) were able to distinguish phonological (language-specific) from phonetic nasalization by manipulating speech rate. They found a quite different distribution of the temporal patterns of nasalization in the two languages, as a function of rate: in American English the extent of nasalization on the vowel preceding the nasal was proportional to the varying vowel duration, while in Spanish it remained constant. They concluded that the spread of nasalization as vowel duration increases in AE must be intentional (phonological), while the short and constant extent of nasalization in Spanish must be an automatic consequence of the speech mechanism, since it reflects the minimum time necessary for the lowering gesture of the velum. However, there is no strict dichotomy between universal and language-specific variations, as coarticulation differs in degree across languages.

References

Borden, G.J., and Harris K.S. 1980. Speech science primer: physiology, acoustics and perception of speech. Baltimore: Williams and Wilkens.

Clark, J., and Yallop C. 1995. An introduction to phonetics and phonology, 2nd edition, Oxford: Blackwell.

Brosnahan, L.F., and Malmberg B. 1970. Introduction to phonetics. Cambridge, England: W. Heffer & Sons Ltd.

Chomsky, N., and Halle M. 1968. The sound pattern of English. New York: Harper & Row.

Davenport, M., and Hannahs S.J. 1998. Introducing Phonetics and Phonology, 196pp. Great Britain: Arnold.

Farnetani, E. 1997. Coarticulation and connected speech processes. The handbook of Phonetic Sciences, ed. by W.J. Hardcastle and John Laver, 371–404. Cambridge, MA: Blackwell.

Ladefoged, P., and Maddieson I. 1996. The sounds of the world's languages. Oxford: Blackwell.

Solé, M.J., and Ohala J. 1991. Differentiating between phonetic and phonological processes: the case of nasalization. Proceedings of the 12 International Congress on Phonetic Sciences, Vol. 2, 110–13, Aix-en-Provence.

ASTRID VAN WIERINGEN

See also **Anatomy of the Articulatory System; Assimilation and Dissimilation; Speech Production**

Assimilation and Dissimilation

All words have citation forms, or dictionary pronunciations, which occur when they are pronounced in isolation. These dictionary pronunciations may undergo sound changes when the word is used in connected speech. One of the most common processes of sound change is assimilation.

'Assimilation' is a cover term that includes a variety of processes and refers to the change in a language sound that makes it more similar to a neighboring sound. The process by which the Latin word *noctem*, meaning 'night', became *notte* in Italian shows an assimilation of the [k] sound, represented by <c> in the spelling, to the same point of articulation as the following [t].

'Dissimilation' is the opposite process, in which sounds become more unlike neighboring sounds. The German word *Kartoffel* 'potato' comes from an earlier form *Tartuffeln*. In this case, the [t] at the beginning of the word became [k], a sound dissimilar to the following [t]. Compared to assimilation, dissimilation is an uncommon process in the world's languages and can be less obvious to an observer. As an example, in many Bantu languages, prefixes must disagree in voicing with the first consonant of the root. Thus, the consonant of the prefix *tu-/du-* for 'I' changes, depending on whether or not the first consonant of the verb is voiced. Thus, *bona* 'to see' starts with a voiced consonant [b] and 'I see' is consequently *tu-bona,* whereas *soma* 'to read' starts with a voiceless consonant [s], and the prefix shows up with a voiced one: *du-soma* 'I read'.

Assimilation can occur both across word boundaries and within words, with sounds affecting each other in three main ways: regressively, progressively, and bidirectionally.

'Regressive' or 'anticipatory' assimilation refers to the effect of a later sound on an earlier sound. One of the most common types of regressive assimilation in many languages involves nasal consonants, which regularly change place of articulation to match the place of articulation of the following consonant. In English, the <n> in the prefixes *in-* and *un-* is usually pronounced [n] only when the following sound is an alveolar consonant, as in *intention.* In *input,* the following sound is bilabial [p]; hence, the <n> is actually pronounced as bilabial [m].

'Progressive' assimilation refers to the effect of an earlier sound on a later one. The English *-s* morphemes found in plural nouns, third person present tense verbs, and possessives are examples of this type of assimilation, being pronounced [s] after voiceless sounds (e.g. *cats, walks, Pete's,* respectively) and [z] after voiced sounds (e.g. *dogs, reads, Mary's*).

'Bidirectional' or 'coalescent' assimilation occurs when two sounds influence each other, as in the common American English spoken forms for an alveolar consonant followed by an unstressed syllable beginning with a [j]. In the following examples, the [t] and [d] sounds join the [j] to create an affricate: *did you* [dɪd juá] ➔ [dɪdʒ u], *can't you* [kænt juá] ➔ [kæn tʃu], *graduate* [gɹæ dʒueɪt], *ritual* [ɹɪtʃu əɫ].

Assimilation especially affects voicing, manner of articulation, and place of articulation. Voicing assimilation is seen in the case of English *-s,* where the voicing of the first consonant carries over to the voicing of the suffix. The effect of assimilation on place of articulation can be seen in the case of English <n>. The overwhelming pattern for this kind of assimilation is for the first consonant to assimilate to the place of articulation of the second, and not the opposite. In English, this occurs commonly with words like *input*, pronounced usually as *imput,* and *income*, frequently pronounced as *iŋ come*. The same *in-* prefix in Latin before the labial sounds [m], [p], and [b] was written as [m], as in *immemorial, impossible, and imbibe*. Evidence of the manner of articulation assimilation with the *in-* prefix can also be seen historically in Latin in words that were borrowed by French and later English. Before the continuant sounds [l] and [ɹ], *in-* was pronounced and written identically to the following sounds, so that *in+legal* and *in+regular* were written as *illegal* and *irregular*.

The most common explanation for assimilation is that speakers adjust the articulation of sounds because it is easier than keeping each sound distinct in the stream of speech. When speakers are producing hundreds of sounds per minute, it is only natural that the precise articulation of sounds will be affected by other sounds near them. There is also evidence that some types of assimilation may have their roots in listener perceptions rather than speaker articulations.

While assimilation takes place in all languages, the extent of assimilation is sensitive to speech rate and register. All other things being equal, assimilation is more likely to occur in casual styles of speech than in more careful, formal styles. The kinds of assimilation and the extent to which it occurs are also relevant in distinguishing differences in varieties of a language. One clear difference between standard southern British and General American English is in the presence or absence of coalescent assimilation in words such as *gradual* or *issue*. The British pronunciation generally does not assimilate the sounds in the middle of words, i.e. the *du* in *gradual* is pronounced [dju], and the *su* in *issue* is pronounced [sju]. The American pronunciation, however, does assimilate, leading to [dʒ u] and [ʃu], respectively.

References

Broe, Michael. 1992. An introduction to feature geometry. In Docherty and Robert.

Docherty, Gerald, and D. Robert Ladd (eds.) 1992. Papers in laboratory phonology II: gesture, segment, and prosody. Cambridge: Cambridge University Press.

Katmaba, Francis. 1989. An introduction to phonology. London: Longman (especially Chapter 5).

Kingston, John, and Mary Beckman (eds.) 1990. Between the grammar and physics of speech. Cambridge: Cambridge University Press.

Lodge, Ken. 1992. Assimilation, deletion paths, and underspecification. Journal of Linguistics 28. 13–52.

Nolan, F. 1992. The descriptive role of segments: evidence from assimilation. In Docherty and Ladd.

Ohala, John. 1990. The phonetics and phonology of aspects of assimilation. In Kingston and Beckman.

JOHN M. LEVIS

See also **Assimilation and Coarticulation**

Austin, John Langshaw

Austin was a leading member of a school of philosophy variously referred to as 'Oxford philosophy', 'linguistic philosophy', or 'ordinary language philosophy' that developed in Britain in the years between the two World Wars and enjoyed its heyday in the late 1940s through the early 1960s. Each of these descriptive terms is, however, arguably inaccurate to some extent. Not all of the members were, for instance, based at Oxford University; nor is it true that all the philosophers at Oxford were sympathetic to its basic tenets. In fact, it may even be argued that the group did not properly constitute a 'school' but at best a loosely knit gathering of philosophers who shared some common views or at least a certain attitude toward philosophy and who regularly gathered for an informal chat for what were famously known as the 'Saturday morning talks'.

Austin published only seven papers and a book during his lifetime. The book was an English translation of *Grundlagen der Arithmetik* (*Foundations of arithmetic*) by the German mathematician and philosopher Gottlob Frege (1848–1925). Austin's major works *How to do things with words* (HDTW*)* and *Sense and sensibilia* (S&S) were both published posthumously. So too was *Philosophical papers*, a collection of 12 papers, most of which had been originally presented at the annual meetings of the Aristotelian Society and published in the conference proceedings. HDTW and S&S are the written versions (based on Austin's own annotations as well as notes prepared by members of the audience) of two series of talks delivered by Austin at Harvard (1955) and the Berkeley Campus of the University of California (1958), respectively.

HDTW is undoubtedly Austin's major philosophical work and the one responsible for his reputation outside the disciplinary bounds of philosophy, mainly linguistics. In it, Austin develops with great verve a philosophical thesis that, despite its apparent simplicity and unpretentiousness, turned out to be a groundbreaking discovery in philosophical analysis. Austin argued that, contrary to what the weight of philosophical and grammatical tradition would have us believe, not all declarative sentences are used to make statements. While a sentence such as 'The cat is on the mat' may indeed be used to make a statement about a certain state of the world involving a feline creature and its whereabouts at the time of speaking, it is not at all clear that words such as 'I hereby pronounce you man and wife' (as uttered by a priest at a given moment during a marriage ceremony) or 'I name this ship the HMS Queen Elizabeth' (as pronounced by, say, the Prime Minister of Britain at the launching ceremony of the latest addition to his country's naval fleet) are used to make statements. These latter utterances do not describe anything in the world; if anything they help alter the world in significant ways by effectively bringing about certain changes. The priest's words, for instance, literally help 'create' a newly wedded couple. Austin coined the term 'performatives' to refer to the utterances of the latter kind that, despite their declarative form, did not result in statements being made. The former, run-of-the-mill declarative sentences were, by contrast referred to as 'constatives'. Constatives, being straightforward declarations, could be assessed in terms of truth and falsity. The performatives were neither true nor false. They could be considered 'felicitous' or otherwise, depending upon a number of attendant circumstances—in the case of the marriage ceremony, for example, the words had to be pronounced by a priest and nobody else, etc. Furthermore, performatives were not descriptive; to utter them was to perform (whereof the neologism) certain acts—later subsumed under the broader category of 'speech acts'.

Pursuing his line of inquiry, Austin painstakingly sought formal criteria by which to distinguish performatives from constatives. Although a number of candidates emerged, such as the use of the first person, present indicative, active voice, etc., counterexamples

soon convinced Austin that the very attempt was a wild goose chase. Even the hypothesis that peformatives were mostly utterances that were culture-specific (such as marrying in a Christian church) turned out be difficult to sustain. This was confirmed by the discovery that even such ordinary utterances such as 'I promise that *p*' (where *p* is any complement denoting a future course of action to be performed by the speaker) meet all the features that help identify performatives. A promise is neither true nor false; it may be described as infelicitous if it turns out, for instance, that the speaker had no intention whatsoever to carry it out. So a promise is as good a speech act as marrying a couple or launching a ship. And so too are such routine acts as ordering, requesting, etc.

Austin was also struck by the fact that a '*p*' simpliciter can just as well do the job as 'I promise that *p*' provided the circumstances are appropriate. Thus 'Do it' as uttered by an army general to a private is readily understood as an order rather than, say, a request, even in the absence of the 'performative formula' 'I order you to …'. The only difference between them is one of explicitness. Accordingly, Austin distinguished between implicit (or 'primary', as he called them) and 'explicit performatives'.

But then, at this precise stage, the Austin reader is only being prepared for a dramatic turn in Austin's reasoning. The sentence 'The cat is on the mat', he argued, is a primary performative whose explicit counterpart would be 'I state that the cat is on the mat.' This means, Austin went on to argue, that there really are no constatives but only some very 'clever' performatives masquerading as such—so successfully indeed that generations of philosophers were tricked by them. Linguistic activity is fully taken up by the performance of speech acts.

The distinction between constatives and performatives was thus replaced by a three-way distinction: locutionary acts, illocutionary acts, and performative acts. Locutionary acts were acts of producing utterances with certain sense, while illocutionary and perlocutionary acts were acts done *in* saying certain things and *by* saying certain things, respectively. The locutionary acts were further subdivided into 'phonetic acts' (acts of merely uttering certain noises), 'phatic acts' (acts of uttering certain vocables or words), and 'rhetic acts' (acts of using the vocables with a certain more or less definite sense and reference). Unlike the illocutionary acts whose performance depends on the speaker saying the words with a certain intention, etc., perlocutionary acts can be said to be felicitous just in case the hearer reacted to the utterance in the way intended by the speaker. Thus, promising is an illocutionary act whereas threatening is a perlocutionary act. The illocutionary force of an act of promise can be made explicit by the use of the performative verb 'promise' as in 'I promise etc.' Curiously, there are no formulas for making explicit the perlocutionary nature of utterances; utterances such as 'I threaten you that …' and 'I persuade you that …' are bizarre, to say the least. This is but a consequence of the fact that the success of a perlocutionary act depends on the hearer doing his/her (i.e. feeling threatened, persuaded, etc.) part rather than the speaker simply intending to achieve a certain goal.

S&S addresses the familiar problem of sensory perception and argues against the then mainstream view patronized by, among others, A.J. Ayer and H.H. Price, to the effect that what one perceives is a sensory impression which may or may not be an accurate representation of the object perceived. Austin argued that the correct answer to the question 'What exactly is one looking at when one gazes a church camouflaged as a barn or a straight stick placed a glass full of water' should be 'A church-camouflaged-as-a-barn and a stick-made-to-look-bent'. In other words, the epistemic state of the observer is part of the objectual reality that is perceived. Now, this is indeed a bold thesis because it flies in the face of the claim made by naïve realists that truth is nonepistemic.

Austin became a household name in disciplines such as linguistics, psychology, sociology, anthropology, etc., thanks to John Searle who took up the torch after Austin's untimely death in 1960. Searle is widely recognized as Austin's intellectual heir. In fact, many textbook presentations of the speech act theory leave the reader with the impression that there is a smooth continuity between the two. Recently, however, the continuity thesis has been challenged by a growing number of scholars. Austin's philosophy continues to attract attention, and many scholars from fields other than philosophy have come up with fascinating readings of his works.

Biography

John Langshaw Austin was White's Professor of Moral Philosophy at the University of Oxford, England (1952–1960). He was born in Lancaster, England on March 26, 1911, and spent his early years in Scotland. At 13, he obtained a scholarship in classics and went to Shrewsbury school. Five years later, he won another scholarship in classics and moved to Balliol College, Oxford. In 1931, he won the prestigious Gaisford Prize for Greek Prose. He moved to Oxford definitively in 1933, when he was examined and elected to a Fellowship at All Souls College. He moved to Magdalen as a tutor in 1935. In 1939, he published his first paper. The following year, he was commissioned in the Intelligence Corps of the Army

and was posted to the war Office in London. In 1941, he married Jeane Courts. In 1944, he moved to Normandy and from there to Versailles. He left the Army in 1945 with the rank of a lieut.-col. and the title of OBE, and was awarded the Croix de Guerre by the French and appointed Officer of the legion of Merit by the Americans. In 1950, he published an English translation of Gottlob Frege's *Grundlagen der Arithmetik* (*Foundations of Arithmetic)*. He visited Harvard in the spring of 1955 where he delivered the William James lectures (later published as HDTW) and the Berkeley campus of the University of California (where he delivered lectures on perception, later published as S&S). He died of cancer in February 1960.

References

Austin, John Langshaw. 1961. Philosophical papers. London: Oxford University Press.

———. 1962. How to do things with words. Oxford: Clarendon Press.

———. 1962. Sense and sensibilia. Oxford: Clarendon Press.

Berlin, Isaiah, et al. (eds.) 1973. Essays on J.L. Austin. Oxford: Clarendon Press.

Cavell, Stanley. 1995. Philosophical passages: Wittgenstein, Emerson, Austin, Derrida. Oxford, UK: Blackwell.

Fann, K.T. (ed.) 1969. Symposium on J.L. Austin. London: Routledge & Kegan Paul.

Felman, Shoshana. 1980. Le Scandale du corps parlant: Don Juan avec Austin ou La séduction en deux langues. Paris: Seuil.

Rajagopalan, K. 2000. On Searle [on Austin] on meaning. Language & Communication 20.

Searle, John R. 1969. Speech acts: an essay in the philosophy of language. New York: Cambridge University Press.

Shaw, H.E. 1990. With reference to Austin. Diacritics 20.

Vendler, Z. 1967. Review of Austin's Sense and sensibilia and How to do things with words. Foundations of Language 3.

Warnock, G. J. 1988. J.L. Austin. London: Routledge.

KANAVILLIL RAJAGOPALAN

See also **Searle, John**

Australia

Australia with its overseas territories is somewhat smaller than the USA and, despite a population of only 20 million, is linguistically richer than many other countries. Although English is its principal language, Australia is host to some 250 other languages as well. It is an interesting area for studies of languages in contact: (1) Many languages have been imported, and English has undergone independent development. (2) Languages and dialects have influenced each other in many ways, with contact languages emerging. (3) Many communities, especially indigenous ones, are bilingual, although many languages have been lost due to overt pressure or a natural shift to English (rarely to another language); however, a number of languages have been reintroduced in the twentieth century, creating a multilingual environment. (4) Australia has developed language and educational policies that have had important effects on the texture of society.

Much of Australia's pre-European human history—the initial immigration, spread, and growth, social organization, and the relationship of indigenous languages with their neighbors—remains obscure. The population was small; most estimates suggest a figure approximately above 300,000. As for the number of languages, experts agree on about 250 languages with 250–300 dialects, which were tied to social units, i.e. tribes, clans, or moieties. (Today, indigenous Australians tend to speak of 'nations' or 'peoples'.) Each language thus averaged 600 speakers, although there were much smaller and much larger ones. But even the larger languages never exceeded a few thousand speakers. According to indigenous creation stories, languages were implanted onto the land that tribes were granted as caretakers, so that the land, people, and language formed a symbolic unit.

No secure hypothesis regarding genetic affiliation with languages outside the continent can be proposed. Although there is a high level of diversity, especially among non-Pama-Nyungan languages, there is evidence for the assumption that all languages have descended from a 'proto-Australian' language. With respect to their typological affiliation there are competing hypotheses. An older one argues that they form a separate group with two families: the non-Pama-Nyungan languages in the central north and northwest and the Pama-Nyungan languages elsewhere. A more recent one maintains that there are clusters of languages definable in terms of descendency, and others that are more related in terms of areal (or contact) features. But all make use of grammatical inflections, mark semantic roles more than syntactic function, have free word order, and have complicated, subtle distinctions in the semantics of pronouns, kinship relations, and so on.

With 250 languages in use by a small population, multilingualism was common and was maintained through social practices: ceremonies, marriage outside the immediate group, and the association of languages with the land. Intertribal communication was facilitated by the fact that neighboring languages shared linguistic features, by the existence of a shared sign language and translators. There is little indication that languages functioned intertribally, although today some are used as lingua francas.

The continent's isolation meant that European colonization began late. The expansion from early settlements—in the Sydney region, Port Arthur (Tasmania), and Norfolk Island—was slow, and not until the twentieth century had the whole continent been explored. Indigenous cultures and languages were thus better maintained in the remote northwest and center. In areas that were settled early, indigenous languages ceased to be used decades ago. Thanks to language policies and community efforts, over 50,000 people today use an indigenous language in addition to English or a contact language. Efforts are under way to revive or document traditional languages and transmit knowledge about them to children so as to halt the breaking of the chain of transmission and their complete loss.

Contact with English has led to borrowing, restructuring, and code switching. Thus, the word *kangaroo* is used in many languages, although it originated in Queensland. The elaborate kinship terminology is gradually being abandoned and English words like *father* and *brother* acquire indigenous shades of meaning. This example from Warlpiri (around Alice Springs) shows code switching: ***September**-rla-rnalu **meeting** wangka-ja, wali natirli yanurnu **committee member**-patu panu purda-nyanjaku.* 'During September, we attempted to hold an AGM [= annual general meeting], but many committee members were unable to attend.' Although a few languages have the vitality to be used in existing domains of communication, most have peripheral status and have lost the association with the land.

The immigrant population reached the 1 million mark as late as the 1850s. Growth accelerated as the herding, fishing, and manufacturing industries became incentives for immigrants. Although immigration has remained a significant factor, natural growth never fell below the 40% mark per decade from the end of the nineteenth century, which ensured that English could develop reasonably independently from British English. But as most immigrants came from Britain and Ireland, there was a continuous influx of native speakers and an outside model to emulate. Its standing was enhanced by the fact that children of well-off parents were educated in Britain and that higher-level staff in the education system, the churches, and later the Australian Broadcasting Corporation were imported from Britain. Speakers of other languages were assimilated or integrated over time and shifted to English without leaving a linguistic trace.

Immigration of non-English speakers has had peaks and troughs. Nineteenth-century migration was fed mainly by Celtic-speaking Britons, Germans, Chinese, and, from the late nineteenth century, Italians. There were small numbers of Jews, Poles, Afghans, Japanese, Malays, and Indians. Asians and Pacific Islanders were brought in to provide labor, as in households or for camel driving, pearl fishing, or sugar cane growing. Gold fields were multilingual sites, but cities became multilingual as well, while indigenous diversity was suppressed. The 'White Australia' policy from the late 1880s to the mid-1960s put pressure on non-whites to assimilate or emigrate. The education acts (Victoria's was the first in 1872), which made education free and secular, as well as World War I and other developments, sped up the loss of immigrant languages, so that Australia assumed an Anglo-Celtic outlook. English became the only medium of communication in the public domain.

The period from 1945 to the 1980s witnessed renewed waves of immigration. Immigrants in the early post-World War II years were predominantly European and reintroduced their 'Languages-Other-Than-English' (LOTEs). The end of the White Australia policy in the late 1960s triggered an influx of Vietnamese, Chinese, and Middle Eastern immigrants who introduced Vietnamese, Chinese languages, Bahasa Indonesia, Arabic, and other languages. Asian languages are today among the strongest LOTEs and have the youngest speech communities. Most European speech communities are aging; only Greek is holding its ground. Approximately 25% of the population use a LOTE 'at home', according to the 2001 census, which includes the 10% of the 450,000 indigenous Australians who use an indigenous language. There are regional differences among states and cities. Sydney has the largest proportion of LOTE speakers and reflects recent Asian and Middle Eastern immigration, while Melbourne remains more European. Language maintenance and shift depend on well-studied factors, but it is worth pointing out differences between different communities speaking particular LOTEs such as Spanish, Arabic, and Chinese languages regarding maintenance and shift. A very significant effect of recent immigration has been the rise of multiculturalism and, more recently, a concept of pluralism and the persistence of a climate of opinion favorable to language policy and LOTE maintenance.

A significant feature is the existence of contact languages. A pidgin was first attested in Sydney as early as the early 1790s. Its spread followed the paths of the expansion of settlement and exploration. The wider

context of trade in the Pacific ensured that it was influenced by the maritime jargon of the Pacific, which itself had been influenced by that in the Atlantic, by local languages and, later, by South Seas jargon and Melanesian pidgin. It was not merely on the receiving end, but it influenced those pidgins during the nineteenth century as well. Kriol and Torres Strait Creole, two offsprings of Australian pidgins, developed in the Northern Territory and north Queensland, respectively, and are used by more than 10,000 speakers each—more than any traditional language. Depending on their 'broadness', they are more similar either to indigenous languages or to Australian English. Due to decreolization, they now participate in a gradient that connects them with Australian English via Aboriginal English.

Aboriginal English is a dialect of Australian English that straddles the space between pidgin and mainstream English, depending on a speaker's background and region. It has replaced indigenous languages as a source of influence on Australian English. Words like *business, sorry (day),* and *shame* reflect indigenous meanings and relate *business* to traditional practices and *shame* to emotions in the presence of strangers. Aboriginal English is increasingly used in literature.

Independent pidgin formation occurred in Western Australia and, possibly, in Tasmania. Norfolk Island Creole and (British) Pitcairnese are worth noting, both of which go back to the early period of colonization. Immigrant ethnic varieties of English have been reported for Italian, Greek, and Asian communities. English, the typical lingua franca in multilingual settings such as the workplace, has acquired discourse features that reflect the broad cultural background of non-English-speaking immigrants.

No language can be understood without taking into account language contact and the wider situation, and this is especially true of English. Today, Australian English is a distinct center of English, but its historical dependency on British English dialects is well known. Its accent and vocabulary have absorbed features from many dialects, especially Cockney and those of southeastern England, Scotland, and Ireland. Australia's demography ensured that dialects from southeastern England provided the basis for Australian English. Vowels in words like *lead, late,* and *loud* are similar to London's Cockney; those in *lid*, *led* are high, as in traditional accents from southeastern England. The vocabulary has retained dialect words, e.g. *tucker,* and words from late eighteenth-century Anglo-American English, e.g. *township, squatter.* One can find traces of Anglo-American English and the northern type of English in the sentence structure of Australian English, as in the definite article in 'he goes to *the* uni' and the avoidance of *shall* outside very formal con-

texts. Australian English is known for its informality, nonstandard grammar, and slang expressions; note expressions like *bullsh* [= bullshit], *bloody* [as an emotive adjective]*, to have kangas in your paddock* 'to be crazy'. As elsewhere in the English-speaking world, nonstandard grammar contains features like double negation; fairly unique to Australian English is the use of *she* for inanimate objects as in *she'll be alright* 'it'll be alright'. Today, Australian English is fully stratified socially but still lacks well-defined regional patterns. For well over a century, Australian English was considered a low derivation from British English, but the last 40 years have witnessed the rise of a standard variety, which is the target of instruction and use in the public domain. It is now well covered in dictionaries, usage guides, and even grammars. The participation of Australian English in international movements, e.g. nonsexist language, plain English, and the dominance of American English are putting pressure on it to adopt international English, especially American English, terms at the expense of local usages.

A crucial feature of Australian English is the influence of indigenous and, to a lesser extent, immigrant languages. Indigenous loans like *boomerang* and *kangaroo* have become international words. The shift in expressions to denote social units is interesting. The term *Aborigine* is being restricted to a generic use, as regional groups have been adopting indigenous names, e.g. *Koori* (southeast) or *Nyunga* (west), which have become a part of standard Australian English. Most loans are, however, little used. Immigrant languages have had little effect, mainly in the sphere of food, although German has contributed a few additional words. As the title of Australia's informal anthem 'Waltzing Matilda' shows, it is often impossible to ascertain whether a word has come into Australian English directly or by some other route. Both *waltzing* and *matilda* are originally German but are unlikely to have been brought in by immigrants.

With about 250 languages in use, the development and implementation of a comprehensive language policy since the late 1980s is a remarkable fact. Its key features are: (1) English is the de facto national language and should be acquired by all citizens; (2) indigenous languages are to be maintained because they link modern Australia to a distant, indigenous past; and (3) immigrant languages complement English for as long as communities wish to maintain them. Some, such as Chinese, Japanese, German, and Spanish, are seen as beneficial to exports, tourism, and so on, and are offered more widely in education. While multilingualism may not continue on the current level, it has widened the consensus on languages for nationally important functions and has shifted the emphasis from 'high cultural' European languages to the ones

that are closer to Australia's geopolitical region. Although pluralism is still official policy, changes in the political climate and financial constraints have reemphasized the role of English.

References

Clyne, Michael. 1991. Community languages. The Australian experience. Cambridge: Cambridge University Press.

Dixon, R.M.W. 2002. Australian languages. Their nature and development. Cambridge: Cambridge University Press.

Encyclopedia of Aboriginal Australia. 1994. Canberra: Aboriginal Studies Press.

Leitner, Gerhard. 1992. Pluricentric English. Pluricentric languages, ed. by Michael Clyne, 178–237. Berlin: Mouton-de Gruyter.

Leitner, Gerhard. 2004a. Australia's many voices. Australian English—the national languages. Berlin: Mouton-de Gruyter.

Leitner, Gerhard. 2004b. Australia's many voices. Ethnic Englishes. Indigenous and migrant languages. Policy and education. Berlin: Mouton-de Gruyter.

Malcolm, Ian. 1995. Language and communication enhancement for two-way communications. Report. Perth: Edith-Cowan University.

Ozolins, Urs. 1993. The politics of language in Australia. Cambridge: Cambridge University Press.

Wurm, Stephen, Peter Mühlhäusler, and Darrel Tryon, (eds.) 1996. Atlas of languages of intercultural communication in the Pacific, Asia, and the Americas. Berlin: Mouton-de Gruyter.

GERHARD LEITNER

See also **English**

Austria

The Republic of Austria is a country situated in Central Europe with an area of 83,853.3 square kilometers and an estimated population of approximately eight million. It is divided politically into nine states (*Bundesländer*) and is bordered by the Czech Republic to the north, Germany to the northwest, Switzerland and Liechtenstein to the west, Italy and Slovenia to the south, and Hungary and Slovakia to the east.

The official language of Austria is German. However, the variety of German spoken in Austria is not identical to that spoken in other areas, such as in the Federal Republic of Germany. Rather, variation is found in the phonological, morphological, syntactical, lexical, and pragmatic levels of language. The most obvious differences between the German in Austria and in Germany are found in the lexicon. Although it is estimated that more than 95% of German vocabulary is common to both these varieties, differences are apparent in everyday vocabulary and also in the vocabulary used in relation to political, social, and economic institutions and structures. Many words specific to Austrian German, so-called Austrianisms, stem from dialects, colloquial speech, or foreign languages. In particular, Austria's proximity to non-German-speaking countries to the south and east is evident in the many borrowed words used in Austria. The Hungarian influence is apparent, for example, in the use of the term *Schinakel* for 'rowing boat', as is the French influence in the use of *Volant* for 'steering wheel'. Differences in the area of syntax include, for instance, the use of the auxiliary *sein* (to be) with the verbs

sitzen (to sit), *stehen* (to stand), *knien* (to kneel), and *hocken* (to squat/ sit) rather than the use of *haben* (to have) as prevalent in Germany. A further example is the use of different genders with particular nouns—while the word for 'salary' is *das Gehalt* in Germany, it is *der Gehalt* in Austria. Variation is also found in the level of language use (pragmatics). It has been established, for instance, that realizations of requests and complaints in Austrian German are longer and include more subjunctives and more modal verbs than in German German. There is no variation in the orthographical level given that the recent spelling reform of the German language was also adopted by Austria.

Up until the early 1980s, the German language was seen from a monocentric point of view. In other words, the German spoken in the Federal Republic of Germany was regarded the standard variety for Austria, as, indeed, it was for other countries in which German is an official language. Any specific features of these subvarieties were consequently viewed as deviations from the main variety. The 1980s witnessed the rejection of such monocentric views and the birth of pluricentricity, whereby German was regarded as a language with a number of national varieties, such as, for example, Austrian German. This view was based on the fact that the German spoken in different countries evidenced linguistic features particular to these areas, and also on the fact that the inhabitants identified with their particular country-specific variety. Current linguistic debate regarding the status of Austrian German concentrates on variants of this

pluricentric thesis. Three particular standpoints are of interest here. Firstly, the pluricentric state-oriented thesis that we have just mentioned suggests that varieties of German should be classified according to the particular nations with which they are associated. This view sees it as irrelevant that features suggested to be specific to Austrian German may also be used in Southern Germany, for example, since such features are not standard in these latter areas—i.e. they are not found in newspapers, for example, while they are in Austria. The pluriareal thesis, on the other hand, suggests that the varieties of German be differentiated according to the areas in which particular features are found, irrespective of national boundaries. Proponents of this view argue that dialects of Southern Germany, such as Bavarian, share many features with Austrian German—consequently, so-called 'Austrianisms' are not Austrian-specific. In addition, they point out that many 'Austrianisms' are not common to Austria as a whole, an east-west divide being identified. The final standpoint is termed pluricentric-neutral and sees Austrian German as the sum of the differences relative to other varieties of German. National or regional factors are irrelevant.

Although the concept of pluricentrism implies that different varieties of a language exist, it does not suggest that all varieties enjoy equal status. The attitudes of the Austrian people to Austrian Standard German vary between viewing it as a soft, melodious language to rejecting it in favor of German Standard German (GSG). Overall, a general, so-called 'linguistic cringe' can be identified among Austrians relative to German Standard German. This stems from the fact that Germany is a vast country, with a much larger population, and much greater economic and political power than Austria. In addition, some semantic confusion exists. Not only does 'German' denote the language but also the people of Germany, leading the 'German' language to be associated with 'the German people'. Also, there is a tendency for Germans to view national varieties of German as regional varieties, given the overlaps that exist between Southern German regional forms and Austrian standard forms. Furthermore, the majority of German language publishing houses are based in Germany, and, thus, commonly demand the use of German Standard German. Finally, the Duden grammar and dictionary, published in Germany, and accepted as a norm in Austria, takes features of German Standard German to be standard and marks Austrianisms in the same way as it does regional variants. Nevertheless, in recent years, an overall increased national awareness has led to a heightened identification with Austrian Standard German. This is reflected, not least, in an increased promotion of the Austrian standard abroad, particularly in the educational context in Eastern Europe.

As far as language planning, i.e. the use of deliberate, often official, measures relating to the status and/or form of a language, is concerned, the codification of Austrian German is yet to be completed. The *Österreichisches Wörterbuch* (ÖWB) (Austrian dictionary), published under the auspices of the Austrian Ministry of Education and the Arts, contains the standard lexicon of Austrian Standard German (ASG) and also includes some grammatical descriptions. However, further detailed descriptions of Austrian Standard German remain to be conducted.

It is not the case that Austrian Standard German is always spoken in Austria. Rather, a complicated continuum between the poles 'dialect' and 'standard' has been identified, with colloquial speech found between the two poles. Regional variation is most pronounced between eastern and western varieties. In addition, due to late industrialization, variation in Austrian German is influenced by social class to a larger extent than is the case in Germany.

Approximately 98% of the population of Austria speak German as their mother tongue. Ethnic minority groups account for the remaining figures. In addition to the ethnic groups of workers and their families from Turkey and the former Yugoslavia who are concentrated in towns and speak Turkish and Serbo-Croatian, respectively, there are four established ethnic minority groups in Austria, namely the Slovenes, the Croats, the Hungarians, and the Czechs, the former two of which are the principal groups. Austrian law ensures the rights of these groups to learn their language in school, to use it in an official capacity, and to mark local signposts bilingually. Speakers are today bilingual, speaking German as well. The coexistence of two or more languages within one state where speakers use these different languages alternately in particular situations (i.e. language contact) has effects both on the development of the languages themselves and also on their status in a particular speech community.

Slovenian is the oldest ethnic minority language of Austria, dating back over a millennium. It is spoken in the southern state of Carinthia (Kärnten) in the Gail, Rosen and Jaun Valley south of Lake Woerth and the capital Klagenfurt, and is divided into several dialects. Language contact has led to changes in the lexical and phonetic levels in both Carinthian Slovenian and Austrian German. *Tscherfl,* borrowed from the Slovenian *črevlje,* is, for example, used in the German Carinthian dialect to mean 'old shoes', and *venahti* from the German *Weihnachten* ('Christmas') is found in Carinthian Slovenian. Despite various efforts by

interested groups to encourage the use of Slovenian, this minority language continues to be used less and less and the majority language, German, is used more and more.

The Croatian minority group has lived in a number of areas in the eastern Austrian state of Burgenland since the fifteenth and sixteenth century at which time whole villages fled the approaching Ottomans and were taken in by feudal lords in need of laborers. These Croatians were speakers of all three Croatian dialects. Today's Burgenland Croatian differs from Croatian proper to such an extent that mutual intelligibility is difficult. This development stems partly from language contact with German and Hungarian, which has led to interference on the level of the lexicon, phonology and syntax, and partly, unlike the case of the Carinthian Slovenians, from efforts of the Austrian Croatian minority to preserve their own identity rather than to nurture links to the homeland—a partial by-product of the fact that whole communities fled together. With the goal of an independent identity in mind, this Croatian minority rejected the adoption of the modern written language of their home country in favor of their own variety of written language, based on Čakaver, the Croatian dialect spoken by the majority of Burgenland Croats. There now also exists a dictionary of Burgenland Croatian. Nonetheless, the trend from monolingualism in Croatian, to bilingualism, and finally to monolingualism in the majority language, German, continues.

Hungarian has been spoken in the state of Burgenland since the tenth and eleventh century when Hungarian guards were sent to the frontier to protect the Hungarian border from the West. This Hungarian-speaking population is concentrated in a number of areas in Burgenland—the most important of which are those in Oberwart (Felsöör) and Oberpullendorf (Felsöpulya). Despite promotion of the language in bilingual schools via the option to learn Hungarian as a noncompulsory subject and the—albeit limited—transmission of Hungarian on television and radio, there is a clear trend toward monolingualism favoring Austrian German.

Finally, Czech is spoken predominantly by former intellectuals of the monarchy in Vienna who, rather than returning to Czechoslovakia upon its new foundation, remained in Austria.

Romany, the language of the gypsies, termed Sinti and Roma by the speakers themselves, originated in India around AD 1000, and is now a language without a particular homeland, not currently spoken in India but rather in Europe and also further afield—in the USA, Australia, and Canada, for example. Corresponding to its wide distribution, there exist approximately two or three dozen dialects of the language, a number of which are currently spoken in Austria. All but exterminated at the hands of the National Socialists in World War II, it is only in recent years that the self-confidence of these speakers has increased, and organizations and interest groups have been established in Austria and elsewhere. In both Burgenland and Vienna, these groups are responsible for the publication of a bilingual newspaper, for example, and once a month a program is also produced and broadcasted by various groups of Romany. Since 1993, the Romany ethnic group has been recognized in Austria. Speakers are usually bilingual, using the minority language predominantly at home and, on occasion, at work also. Dialects of Romany have only been codified to a small extent—the dialect of Romany spoken in the Austrian state of Burgenland being one example of a codified variety.

The principal foreign languages on offer in Austrian secondary schools include English and French, and occasionally Italian or Russian. Latin and Classical Greek are also taught.

References

Ammon, Ulrich. 1995. Die deutsche Sprache in Deutschland, Österreich und der Schweiz. Das Problem der nationalen Varietäten. Berlin and New York: de Gruyter.

Barbour, Stephen, and Patrick Stevenson, 1990. Variation in German. A critical approach to German sociolinguistics. Cambridge: Cambridge University Press.

Clyne, Michael G. (ed.) 1992. Pluricentric languages. Differing norms in different nations. Berlin and New York: de Gruyter.

Clyne, Michael G. 1995. The German language in a changing Europe. Cambridge and New York: Cambridge University Press.

Martin, Victoria. 2000. The German language in Austria. National varieties of German outside Germany, ed. by Hogan-Brun, Gabrielle, 103–19. Bern: Lang.

Muhr, Rudolf, and Richard Schrodt, (eds.) 1997. Österreichisches Deutsch und andere nationale Varietäten plurizentrischer Sprachen in Europa. Empirische Analysen. Vienna: Hölder-Pichler-Tempsky.

Ohnheiser, Ingeborg, Manfred Kienpointner, and Helmut Kalb (eds.) 1999. Sprachen in Europa: Sprachsituation und Sprachpolitik in europäischen Ländern. Innsbruck: Institut für Sprachwissenschaft.

Österreichisches Wörterbuch (ÖWB), 35th edition, 1979. Vienna: ÖBV and Jugend & Volk; 36th revised edition, 1985; 37th revised edition, 1990; 38th revised edition, 1995.

Russ, Charles V.J. (ed.) 1990. The dialects of modern German: a linguistic survey. London: Routledge.

Wiesinger, Peter (ed.) 1988. Das österreichische Deutsch. Vienna: Böhlau.

ANNE BARRON

See also **Germany; Language: Contact—Overview; Language Death; Language Planning; Standard Language**

Austroasiatic

The Austroasiatic language family contains languages spoken by about 80 million people, primarily in Southeast Asia; however, only two of the languages in the family, Vietnamese and Khmer, have national status. Nowhere do the speakers of major Austroasiatic languages appear to be recent migrants. This is a language family that must have once filled more of the map of Southeast Asia and eastern India, now reduced to scattered islands by encroaching Indo-Aryan, Sino-Tibetan, Tai, and Austronesian languages, including Bengali, Assamese, Burmese, Thai, Lao, and Cham. The Austroasiatic language family is thought to have two major branches, which are typologically very divergent from each other. The division between the Mon-Khmer group and the Munda languages must have taken place many thousands of years ago.

The Munda or Western branch of the Austroasiatic family, spoken entirely in east central India, is distantly related to Mon-Khmer. It shares the South Asian subcontinent with three other language families. Indo-European languages, particularly the Indo-Aryan branch of Indo-European, dominate in the north, while the south is the domain of the Dravidian languages. The subcontinent's northern fringe is occupied by Sino-Tibetan languages, and Austroasiatic languages are scattered from central India eastward into Vietnam. In India itself, the Austroasiatic language with the most speakers is Santali, but the major languages of this family, Vietnamese and Khmer (Cambodian), are spoken in Southeast Asia. The major Sino-Tibetan languages are spoken in East and Southeast Asia (e.g. Chinese, Burmese). In addition to these four families, there is one language isolate, Burushaski, spoken in northern Pakistan, while the genetic affiliations of the Andaman Islands remain unclear.

The Munda branch consists of two groups or sub-branches. The Nahali or Nihali sub-branch has about 5,000 speakers in Madhya Pradesh and Maharashtra, northeastern India, and the genetic affiliation of this language is controversial. The Munda sub-branch is found in northeastern India, in the mountains and plateaus of Madhya Pradesh, Bihar, and Orissa. It is geographically surrounded by Indo-European and Dravidian speakers, whose languages have greater social prestige. It consists of a fairly large number of languages, including Santali (spoken mainly in India, but also in Bangladesh and Nepal), Ho (spoken mainly in Bihar and Orissa, India), Sora (spoken mainly in Orissa, India), Korku

(spoken mainly in Madhya Pradesh, India), Juang (spoken mainly in Orissa, India), Korwa (spoken mainly in Bihar and Madhya Pradesh, India), and Mundari (spoken primarily in Assam and Bihar, India). Mundari are typically hill peoples, while the related Santali is spoken by lowland farmers. Mundari and other Munda languages have been spoken alongside Indo-Aryan languages for two millennia. They show the signs of this long coexistence not only in loan words but also in the structure of typical sentences. Santali is traditionally an unwritten language; speakers who learn to read and write do so in Bengali or Oriya, and on the occasions when Santali is written, Bengali or Oriya scripts are generally used. As with its close relative Mundari, Santali and its speakers have little recorded history before the nineteenth century.

Munda languages appear to have been greatly influenced by their Indo-European and Dravidian neighbors. In contrast to these neighbors, Munda languages have traditionally been unwritten, and their peoples have little recorded history until they came into the sights of British administrators in India's colonial period. Typologically, Munda languages are agglutinative, and very long sequences of affixes may be found, especially in verbs. There is even noun incorporation (e.g. in Sora), which is usually a feature of polysynthetic/incorporating languages. There are two genders for nouns in most of the Munda tongues: animate and inanimate. Munda languages use affixes extensively. Case relations are signaled by word order, postpositions, and pronominal affixes. The basic word order is subject–object–verb. Most Munda languages also have three numbers—singular, dual, and plural. Suffixes and particles placed after the noun are used to express such features as number and possession, which are often indicated in Indo-European tongues by case inflection.

The Mon-Khmer or Eastern are typologically quite different from the Munda languages. The Mon-Khmer languages are mostly prefixing, monosyllabic, usually with a large number of vowel contrasts. The most commonly found word order is subject–verb–object. Vietnamese has borrowed extensively from Chinese, especially during the Tang dynasty, when Vietnam came under the strong cultural and political influence of China, and was even written with Chinese characters until Western Christian missionaries introduced the Roman alphabet. Grammatically, the Mon-Khmer

languages make considerable use of affixes (prefixes, infixes, and suffixes). They are agglutinative in that different linguistic elements, each of which exists separately and has a fixed meaning, are often joined to form one word. Cambodian and Mon have their own scripts, which descended from the alphabets of India. Both are written from left to right.

Scholars have long realized that Khmer and Mon were related, and Mon-Khmer is the well-established group of Austroasiatic languages typified by these two. Mon language is now in decline, as Burmese and Thai expansion supplanted Mon rule. Through Mon, Buddhist culture was transmitted to the early speakers of Burmese. By way of Mon script, Burmese and Shan first became written languages. Mon is now a minority language of southern Burma and central Thailand, and most of its speakers are bilingual in Burmese or Thai. Although Mon is not an official language in Burma, it is now being used again in traditional monastic education, which most boys receive. Mon is not a tonal language but, like Khmer, has a 'clear' and a 'breathy' register, a feature that now forms part of the Burmese sound pattern as well. Modern Mon can be divided into three dialect groups: Pegu, Martaban and Moulmein, and Ye. Mon villages in central Thailand, resulting from settlements in the last three or four hundred years, speak dialects close to those of Martaban and Moulmein.

Khmer language survives, and is now the majority language of Cambodia. Close to Khmer are some minority and hill languages of Southeast Asia, including Sre, Mnong, Stieng, Bahnar, Hre, Sedang, Kuy, Bru, So, and others. Khmer script, like others of South and Southeast Asia, descends from the Brahmi of ancient India.

Mon-Khmer consists of six sub-branches: Nicobarese, Northern Mon-Khmer, Eastern Mon-Khmer, Viet-Muong or Vietic, Monic, and Aslian. Nicobarese consists of six different languages with a total of over 22,000 speakers on Nicobar Islands (under Indian administration) in the Bay of Bengal. These include Car, Nancowry, Great Nicobarese, and others. In these languages, taboo leads to word avoidance on a considerable scale, and thus to a rapid turnover of vocabulary. Thus, despite their dissimilarities among each other, and with other Mon-Khmer languages, the relationships are historically closer than this would suggest.

The Northern Mon-Khmer sub-branch has many languages of very small communities, including Mlabri, spoken by 300 hunter-gatherers. This sub-branch has three sub-branches itself: Khasi or Khasian, Palaungic or Palaung-Wa, and Khmuic. Khmuic consists of several languages spoken in northern Laos and northern Thailand, of which the main one is Khmu.

Khasi or Khasian is spoken mainly in Assam, northern India, and Bangladesh, with over 500,000 speakers. Khasi is the state language of the Indian state of Meghalaya, whose capital is Shillong, and it has many loan words from Bengali, Urdu, and English.

Palaung-Wa, formed by two groups, Wa and Palaung, consists of a large number of minor languages spoken in scattered pockets located in Thailand, Laos, Burma, and Yunnan. Parauk is one of the more prominent languages in this group. Most of the 650,000 speakers live in Burma and China. Wa probably has well over 1,000,000 speakers. The majority of speakers live on the mountainous borderland between Burma and China. The legendary center of the Wa country is Lake Nawngkhio, high in the mountains on the China–Burma frontier. The older history of the language and its speakers is unknown, but Austroasiatic speech may well have a history of several millennia in this region. Apart from their mountain heartland, Wa speakers are found in scattered communities in mountainous parts of Xishuangbanna, the southern Shan State, and northwestern Thailand. Wa languages are largely monosyllabic; they have no tones, but tend to have rich and complicated vowel systems.

The Eastern Mon-Khmer sub-branch is further subdivided into Khmeric, Bahnaric, and Katuic spoken in Vietnam, Malaysia, Thailand, Cambodia, and Laos. Khmeric consists of a single language, Khmer. Most of the over 7 million speakers reside in Cambodia, with others in Vietnam, Malaysia, and Thailand. The Bahnaric sub-branch includes about 35 minor languages spoken in central and southern Vietnam, southern Laos, and eastern Cambodia. Bahnar, with over 85,000 speakers in Vietnam, is one of the major languages. And Katuic sub-branch includes several minor languages scattered in Vietnam, Laos, Thailand, and Cambodia. The Monic sub-branch includes Mon and Nyahkur, spoken in Burma and Thailand. And, the Aslian languages (roughly 40,000 speakers) of inaccessible districts in the Malay Peninsula and southern Thailand include Kensiu, Sengoi, Orang Benuja, Temiar or Northern Sakai, and others with even small numbers.

The Viet-Muong or Vietic sub-branch includes Vietnamese and a number of minor languages spoken in northern Vietnam and northern Laos. Most Vietnamese speakers reside in Vietnam, with the rest in the United States, Cambodia, and China. Muong is spoken in inland northern Vietnam. It has five tones and, like Vietnamese, is a monosyllabic language. It is likely that the speakers of early Muong and of early Vietnamese lived not far apart 2,000 years ago.

The inclusion of Vietnamese and Muong (with some tiny minority languages of Vietnam and Laos) in

the Austroasiatic family continues to be seen as controversial by scholars, since Vietnamese has been under the influence of Chinese for 2,000 years. Whatever its shape at the beginning of this period, Vietnamese is now a tonal language (with six tones that frequently help to distinguish homonyms): its sound pattern rather resembles that of Chinese. Vietnamese is basically monosyllabic, but it has many words of two or more syllables. It uses particles but has no prefixes and suffixes. Word order is very important for showing grammatical relationships since there is no inflection. The vocabulary has many loan words from Chinese; Chinese words now make up as much as 60% of the vocabulary of written Vietnamese. Many of these loans arrived before the tenth century, as is evident when their Vietnamese pronunciation is compared with modern Chinese. An alphabet based on Roman letters and adapted for Vietnamese, as by adding diacriticals, is generally used today in place of the traditional Chinese-type writing of the past.

According to another school of thought, the Mon-Khmer branch is actually two branches. In other words, the Austroasiatic language family is believed to have three branches: the Mon-Khmer languages, the Munda languages, and the Annamese-Muong branches. There is considerable evidence but no definite proof as yet that these groups are derived from a single ancestor language, which is the essential requirement for classification in the same linguistic family. The Annamese-Muong subfamily is composed of Muong and Vietnamese (also called Annamese). The classification of Vietnamese is still disputed; some regard it as a Mon-Khmer tongue, others as a Tai (or Thai) language, and still others as a language unrelated to any other known tongue. It is now generally accepted by nearly all specialists that Vietnamese resembles its Austroasiatic neighbors.

On the basis of some lexical and typological similarities, some linguists speculate that Austroasiatic languages are very distant relatives of Sino-Tibetan. Also, Schmidt (1906) first proposed a relationship between Austroasiatic and Austronesian (the so-called Austric hypothesis) on the basis of certain phonetic, lexical, and grammatical similarities, which was discussed further by Benedict (1973) and Headley (1973). However, most linguists currently maintain that there is very little solid, convincing evidence so far to support either of these hypotheses. The single Austric family grouping has not yet been generally accepted.

References

Benedict, P.K. 1973. Austro-thai and Austroasiatic, the 'Austro Linkage'. Paper presented at the First International Conference on Austroasiatic Linguistics, Honolulu, January 2–6, 1973.

Headley, R.K., Jr. 1973. Some sources of Chamic vocabulary. Paper presented at the First International Conference on Austroasiatic Linguistics, Honolulu, January 2–6, 1973.

Henderson, E.J.A. 1991. Austroasiatic languages: essays in honour of H.L. Shorto. London: School of Oriental and African Studies.

Nguyen, D.-H. 1987. Vietnamese. The world's major languages, ed. by B. Comrie, 777–96. New York: Oxford University Press.

Schmidt, W. 1906. Die Mon-Khmer Völker. Ein Bindeglied zwischen Völkern Zentral-Asiens und Austronesiens. Archiv für Antrhopologie 5. 59–109.

Zide, N.H. (ed.) 1966. Studies in comparative Austroasiatic linguistics.

CAMILLA COCKERTON

See also **Indo-European 1: Overview**

Austronesian

The Austronesian language family (formerly Malayo-Polynesian) is the most widespread family in the world and the largest in terms of number of languages, with approximately 1,200 (one fifth of the world's total). The languages are spoken by more than 270 million people in the Malay Peninsula, Madagascar, Indonesia, the Philippines, the island groups of the Central and South Pacific (except for Australia and much of the interior of New Guinea), and New Zealand, as well as in scattered areas of Vietnam, Cambodia, Laos, and Taiwan. Although both the Indo-European and Sino-Tibetan families have considerably more speakers, Austronesian includes some of the world's largest languages in terms of number of speakers, as well as some of the smallest.

Indonesia has 13 Austronesian languages with a million or more speakers, including Indonesian with 17 million, Javanese with 75 million, Madurese with 13 million, Sundanese with 27 million, Balinese and Buginese with 4 million, and Acehnese with 3 million.

Javanese alone accounts for about one quarter of all speakers of Austronesian languages and has more speakers than all of the languages in the Oceanic group.

The other large languages are found in the Philippines, all of whose 160-some languages are Austronesian. Eight have more than a million speakers, such as Tagalog, which is the official language but also one of many regional languages spoken on hundreds of islands with 17 million speakers, Cebuano with 15 million, and Ilocano with 8 million, Hiligaynon with 7 million, Bikol and Waray with 3 million, and Kapampangan and Pangasinan with 2 million.

Outside the Philippines and Indonesia, there are only two Austronesian languages with more than a million speakers: Malay with 17 million and Malagasy on Madagascar, an island off the southeast coast of Africa, with 9 million. Malagasy is the westernmost member, and although it has some influence from mainland African languages, it is classified as Austronesian on the basis of its predominantly Austronesian vocabulary and structure.

Fourteen of the 21 or 22 Austronesian languages spoken by the pre-Chinese aboriginal population on Taiwan are still spoken, although they have been largely submerged by the influence of Chinese immigration from the mainland. The largest language has only 100,000 speakers, and most of the others have only a few thousand or fewer. Some languages, such as Siraya, are already extinct and are documented only in historical texts.

Other Austronesian languages include the Polynesian languages, such as Tongan, Tahitian, and Samoan, all of which have relatively few speakers, particularly Hawaiian, which is severely endangered. In Vanuatu, none of the indigenous Austronesian languages has more than 3,000 speakers, and some there (e.g. Lakona) as well as in New Guinea (e.g. Sirak) have as few as (or fewer than) several hundred speakers. Approximately 600 Austronesian languages are spoken in New Guinea and in scattered coastal areas and on the islands to the east of it, but there are probably fewer than 2 million speakers.

Most of the well-described languages are the larger ones in the Malay and Indonesian archipelagoes and in the Philippines. Politically, the most important language is probably Malay, native to the Malay Peninsula, adjacent portions of southern and central Sumatra, and some smaller neighboring islands. The oldest attestations of the language (in fact, the oldest of any Austronesian language) are found on stone inscriptions in the late seventh century in southern Sumatra. After Islam was introduced at the end of the thirteenth century, Malay-speaking sultanates were established in Malay-speaking regions of the Malay Peninsula, as well as in Brunei on the coast of northwestern Borneo. Malay was spoken on both sides of the strategic Strait of Malacca between Sumatra and the Malay Peninsula. It became an important trade language in the larger region when the India–China trade began around the start of the first century because the preferred route passed through the Strait of Malacca. When the Dutch East India Company arrived in Indonesia at the beginning of the seventeenth century, they found Malay in use as a lingua franca in major ports throughout the archipelago, a role it has retained to some extent today. Malay was chosen as the basis for the national language of Malaysia (Bahasa Malaysia), Brunei (Bahasa Kebangsaan 'national language'), and Indonesia (Bahasa Indonesia), where Malay was felt to be a more neutral alternative to Javanese, the most widely spoken language.

The Austronesian language family is approximately 6,000 years old and can attribute its present wide distribution to the development of open-ocean travel by the ancestors of the present-day speakers. With its multitude of languages on thousands of islands, it has been of key importance in the development of the comparative method in historical linguistics, as well as critical in the study of what clues languages can give us about where the homeland of the language family must have been and what kind of culture the speakers of the ancestral proto-language must have had. Scholars have used the spread of Austronesian to look at issues of importance in human history, such as the spread of agriculture, which Peter Bellwood (1991) argues to be the driving force behind the spread of this language family and others.

Both linguistic and archeological evidence indicate an initial dispersal of Austronesian languages from Taiwan several centuries after Neolithic settlers introduced grain agriculture, pottery making, and domesticated animals to the island from the adjacent mainland of China, approximately 4000–6000 BCE. The linguistic evidence suggests a steady southward and eastward movement, with Austronesian speakers moving around the northern coast of New Guinea into the western Pacific in approximately 2000 BCE. From New Guinea and the Bismarck Archipelago, settlers spread out very rapidly, crossing the ocean in outrigger canoes. There is, however, a puzzling thousand-year gap before the settlement of central and eastern Polynesia, with Hawaii settled only within the past 1,500–1,700 years and New Zealand within roughly the past millennium. Madagascar was settled by immigrants from southeastern Borneo sometime between the seventh and thirteenth centuries CE.

An early split separated the aboriginal Austronesian languages of Taiwan (Formosan) and the rest of the family (Malayo-Polynesian). The designation

'Malayo-Polynesian', formerly applied to the language family as a whole, has been restricted since the mid-1960s to the non-Formosan subgroups of Austronesian. Formosan, however, does not represent a subgroup defined by exclusively shared linguistic innovations. Rather, it is a collective term for a highly diverse group of languages, perhaps as many as six, most of which share broad typological similarities with languages in the Philippines and some other areas, such as Madagascar. Western Malayo-Polynesian is a cover term for the Austronesian languages of the Philippines, western Indonesia (Borneo, Sumatra, Java-Bali-Lombok, Sulawesi), mainland Southeast Asia,

Samoan, and Tahitian, which are remarkable for their wide geographic spread yet close relationship. In addition, some 18 so-called Polynesian Outliers are found in Micronesia and Melanesia. Yapese also seems to be Oceanic, although its place within this group remains uncertain, partly because it shares little vocabulary with more typical languages because of a complex history of borrowing.

The list below shows some very basic items of vocabulary, which reveal the general similarities among some of the languages that have allowed linguists to reconstruct the Austronesian family and to analyze its internal relationships.

Cognates for	'2'	'5' (also 'hand')	'fish'	'fire'	'me/I'
Indonesian	*dua*	*lima*	*ikan*	*api*	*saya/aku*
Malagasy	*roa*	*dimy*	*tsondro*	*afo*	*izaho/ahy*
Tagalog	*dalawa*	*lima*	*isda*	*apoy*	*ako*
Fijian	*rua*	*lima*	*ika*	*buka*	*au/noqu*
Hawaiian	*lua*	*lima*	*i'a*	*ahi*	*a'u/(w)au*

Madagascar, and at least Chamorro and Palauan in western Micronesia. These languages do not exhibit any of the innovations characteristic of Central-Eastern Malayo-Polynesian. They include some of the largest and best-known Austronesian languages, including Malay, Javanese, Acehnese, Sundanese, Balinese, Buginese, Ilokano, Tagalog, Cebuano, and Malagasy.

The Central Malayo-Polynesian languages are found throughout much of eastern Indonesia, including the Lesser Sunda Islands from Sumbawa through Timor and most of the Moluccas. Many of the changes that define this subgroup apply to most of the languages but do not reach the geographic extremes, leading some scholars to question it. The small group of languages comprising the South Halmahera-West New Guinea subgroup is found in the northern Moluccan island of Halmahera and in the Doberai Peninsula (also called Vogelkop or Bird's Head) of western New Guinea. Most of the languages are known only from short word lists, although preliminary descriptions exist for Buli of Halmahera and Numfor-Biak and Waropen of western New Guinea.

The Oceanic subgroup is the largest and best defined of all major subgroups in Austronesian, including all the languages of Polynesia, all the languages of Micronesia (except Palauan and Chamorro), and all the Austronesian languages of Melanesia east of the Mamberamo River in Irian Jaya. Some of the better-known languages are Motu of southeastern New Guinea, Tolai of New Britain, Mota of the Banks Islands in northern Vanuatu, Chuukese of Chuuk in Micronesia, Fijian, and the many Polynesian languages, including Maori, Hawaiian, Tongan, Niuean,

The Austronesian languages have some of the largest and smallest sound inventories, with these inventories showing areal characteristics. Hyenghène in New Caledonia has 30 consonants, whereas Hawaiian has only eight. Sakao in Vanuatu has 12 vowels, whereas several languages of the Philippines and western Indonesia, such as Cebuano Bisayan and Banjarese, have only three. Rare sounds include pre-nasalized bilabial trills (made by trilling the lips after [m]) of the languages of Manus Island and apico-labial sounds (made by touching the upper lip with the tip of the tongue) of some central Vanuatu languages. Some languages, such as Kiput in Borneo, distinguish short and long vowels as well as short and long consonants. Many Oceanic languages permit no final consonants or consonant clusters. Tonal contrasts are also found in a few Austronesian languages, e.g. in the Chamic languages as well as in two widely separated parts of New Guinea and in southern New Caledonia.

Languages with the word order verb–subject–object (VSO) and verb–object–subject (VOS) are common in Taiwan, the Philippines, and northern Borneo. Other VSO languages are Malagasy and Old Javanese. Although the Polynesian languages are also VSO, they have developed this order through a prior subject–verb–object (SVO) stage, unlike the other languages, which have inherited it from Proto-Austronesian. SVO languages are widely distributed in western and eastern Indonesia, Melanesia, and Micronesia. Subject–object–verb (SOV) languages are restricted to the New Guinea mainland and some nearby islands, where they have been in contact with characteristically SOV Papuan languages.

Most languages of the Philippines and Taiwan, as well as the languages of northern Borneo, northern Sulawesi, and Chamorro and Malagasy, mark verbs with affixes indicating the grammatical role that a specific noun bears in relation to the verb. One of the most fundamental distinctions in the verb systems of Austronesian languages is the division into stative and dynamic verbs. Stative verbs often translate as adjectives in English (e.g. Hawaiian *make* 'die/to be dead'), and in many Austronesian languages it is doubtful whether a category of true adjectives exists.

Most of the languages distinguish between inclusive first-person plural pronouns ('we', listener included) and an exclusive form ('we', listener excluded). Some Oceanic languages distinguish a dual (i.e. 'we/you/they two'). The distinction between alienable and inalienable possession is also characteristic of Oceanic languages. Nouns referring to inalienably possessed entities (e.g. body parts, relatives) are typically a small, closed set, and the nouns referring to alienable possessions can be freely added to. Inalienable and alienable possession are often manifested grammatically in different construction types, and some languages have several subclasses of alienable possession, each with its own marker. Other languages, including many in Micronesia, have developed a complex system of classifiers to mark categories of alienable possession. In Pohnpeian, alienably possessed nouns belong to one of more than 20 different classes, whose members usually have some distinctive semantic feature.

Historical connections are complicated by extensive long-term contacts and movements between Austronesian- and non-Austronesian-speaking peoples, which has also brought about drastic grammatical convergence as well as extensive borrowing of vocabulary. There is still some dispute over the membership of many languages. The Maisin language of southeastern Papua New Guinea is now generally regarded as an Austronesian language with heavy contact influence from Papuan languages. Other controversial or aberrant languages include those of Buka and Bougainville, Arove, Lamogai, and Kaulong of New Britain in Papua New Guinea, Ririo and some other languages of the western Solomons, Asumboa of the Santa Cruz archipelago, Aneityum and some other languages of southern Vanuatu, several languages of New Caledonia, and Nengone and Dehu of the Loyalty Islands in New Caledonia. Most of the Chamic languages distinguish words with different intonations and have also acquired other features characteristic of the Mon-Khmer languages with which they have been in contact in Vietnam and southern China.

References

Bellwood, Peter. 1991. The Austronesian dispersal and the origin of languages. Scientific American.

Blust, Robert A. 1995. The prehistory of the Austronesian-speaking peoples. Journal of World Prehistory I: 4.

Dutton, Tom, and Darrell Tryon (eds.) 1994. Language contact and change in the Austronesian world. Berlin: Mouton de Gruyter.

Lynch, John. 1998. Pacific languages. Honolulu: University of Hawaii Press.

Tryon, Darrell T. 1995. Comparative Austronesian dictionary. An introduction to Austronesian studies, 4 vols. Berlin: Mouton de Gruyter.

SUZANNE ROMAINE

Auxiliaries

Auxiliaries are a special type of verb typically expressing the grammatical notions of tense (e.g. past or present), aspect (e.g. finished or continuing), and mood (i.e. commitment to the truth of the utterance). Auxiliaries do not occur by themselves but always 'help' other verbs, hence their name auxiliary, which can be translated as 'helping verb'. Examples of English auxiliaries are *have*, as in *I have eaten*, and *might*, as in *It might rain*. Examples of verbs that are not auxiliaries are *surmise*, *perceive*, *swim*, *understand,* and *eat*. These lexical verbs carry more meaning than auxiliary verbs and are referred to as main verbs. They can occur by themselves, without auxiliaries, as in *they understood that problem*.

The number of auxiliary verbs is limited and they can comprise a relatively short list (about 13 in English). They are often referred to as function words, which are considered a closed class. On the contrary, the number of lexical verbs is vast, and new verbs are invented daily, e.g. *to bookmark a URL*. Hence, they are members of the open class or lexical words.

In addition to lacking meaning, auxiliaries possess a number of unique characteristics of word formation and syntax. For instance, they tend to be unstressed

and therefore can be reduced, or contracted, as in *I'm going there* and *I've seen him*. They are also positioned differently. In English, auxiliaries occur in questions (namely before the subject, as in (1)) and in negative sentences (namely before *not/n't*, as in (2)). If an auxiliary is not present, main verbs need the inclusion of a form of *do*, as in (3) and (4), to question or negate sentences:

(1) Has he eaten yet?
(2) He hasn't eaten yet.
(3) Did he eat yet?
(4) He didn't eat yet.

In other languages, auxiliaries also occupy special positions in the sentence and are marked for negation. In English, auxiliaries are used in tag questions, those tagged onto the end of a statement, such as, *He has eaten, hasn't he*. Again, if a sentence only contains a main verb, the inclusion of a form of *do* is needed, as in *He went, didn't he?*

In English, the auxiliary verbs include (a) *have, be, do, get* and (b) *may, might, can, could, shall, should, will, would,* and *must.* Those listed under (a) can also be used as main verbs, as in *He has a headache*, and the ones listed as (b) are called 'modals', since they help to express mood and modality. A few auxiliaries are marginal since they have characteristics of both main verbs and auxiliaries. For instance, *dare* occurs both with and without *do* in questions, as in (5) and (6), and in negations:

(5) Shall I part my hair behind? Do I dare to eat a peach? (T.S. Eliot in The Love Song of J.A.P. 122)
(6) Dare I eat a peach?

Other marginal auxiliaries (also called semimodals or semiauxiliaries) include *ought (to), need (to), have (got) to, be able to, be going to, be about to, be to,* and *used to*. Modern English is quite unique in being able to 'stack' the auxiliaries, as in (7):

(7) She may have been being seen (while committing that murder).

Historically, auxiliaries have evolved from main verbs. The discovery of these gradual changes inspired many linguistic theories on the nature of linguistic change. Thus, *I have to read a book* is seen as having developed from possessive *I have a book (to read)* to purposive *I have a book to read* to obligation *I have to read a book*. Originally, *have* was a main verb but it developed into a modal. *Have to* is moving toward the reduced form *hafta* in many dialects, but does not yet behave like a modal in terms of its position in negative sentences and questions.

In the late 1970s and early 1980s, auxiliaries cause a very lively debate as to whether the category of AUX, introduced in Chomsky's Syntactic Structures, is universal. Since then, some linguists have argued that all auxiliaries are simply verbs, and others that they are special and that this special category exists universally, hence the similarities between languages as diverse as Japanese, Arabic, and Lummi (a language spoken in the NW of the US). Yet others, working mainly within cognitive linguistics, see auxiliaries in terms of a continuum from auxiliary to main verb, making it possible to think of aspectual verbs, such as *continue, go,* and *got* in *She continued writing, She went fishing,* and *He got going,* as somewhat auxiliary-like and somewhat verb-like. The latter approaches are also interested in what kinds of verbs are the sources of auxiliaries.

Auxiliaries form part of a language's Tense, Mood, Aspect (TMA) system. Even though the core auxiliary is said to express TMA, they may also express passive voice and negation. TMA represents a huge collection of different notions, and not all languages express all. English auxiliaries indicate progressive aspect (as in *I am reading*), present relevance (*I have lived here a long time*), voice (*It was read*), uncertain mood (*It might rain*), polite mood (*Could you please do that*), ability and volition, also considered mood. Other languages indicate how the speaker acquires the knowledge (by direct perception or via hearsay). The similarities of these systems in languages that have developed independently of each other has prompted much debate as to whether this category is biologically innate, i.e. part of Universal Grammar.

In short, auxiliaries are different from main verbs in terms of position, unstressed pronunciation, and uses as TMA markers. This has prompted many linguists to consider them separate from main verbs, and to regard this as a universal property of languages.

References

Chomsky, Noam. 1957. Syntactic structures. Den Haag: Mouton.

Heine, Berndt. 1993. Auxiliaries. Oxford: Oxford University Press.

Krug, Manfred. 2000. Emerging English modals. Berlin: Mouton-de Gruyter.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. A comprehensive grammar of the English language. London: Longman.

Steele, Susan, Adrian Akmajian, Richard Demers, Eloise Jelinek, Chisato Kitagawa, Richard Oehrle, and Thomas Wasow. 1981. An encyclopedia of AUX: a study of cross-linguistic equivalence. Cambridge; MA: MIT Press.

Warner, Anthony. 1993. English auxiliaries. Cambridge: Cambridge University Press.

ELLY VAN GELDEREN

See also **Chomsky, Noam; English**

B

Babylonian Traditional Grammar

Ancient Mesopotamia (Iraq) witnessed the first attempts at studying the structure of language. These attempts are preserved in grammatical texts dating from Old Babylonian times (2000–1600 BC), more than a thousand years before Pāṇini's grammar of Sanskrit. As in India and Europe, grammatical study in Mesopotamia emerged following major concerns with the preservation of a classical literature written in a language that was disappearing at the time.

In the Old Babylonian period, two languages were spoken in the area: Akkadian, a group of Semitic dialects (with Babylonian as a principal division), and Sumerian, an isolated language among them. Sumerian, the language of a major part of Ancient Mesopotamian literature, was dying out as a living language and was being replaced by Akkadian. Nevertheless, Sumerian was to remain the literary language used in administrative, legal, scholarly, and religious areas. Texts written in cuneiform script appeared for the first time around 2900 BC. They continued to be written until the beginning of the Christian era.

Since Akkadian coexisted with Sumerian, the awareness of linguistic diversity was alive among their speakers. Sumerian was taught at school and the cuneiform script was the first type of writing that students learned. Copying texts at school was a duty, writing (additional) lines being used as a punishment. Sumerian grammatical texts were subject matters on exams, together with mathematical problems. It is believed that Sumerian grammar was still being taught in southern Mesopotamia under its Persian rulers about a decade before the battle of Marathon (490 BC) (cf. Black 1989:76).

The intricacy of the cuneiform writing system, as well as Sumerian language and culture were taught by means of lists of signs, bilingual vocabularies, and proper grammatical texts. As education in ancient Mesopotamia was orally conducted and it mainly involved memorization, these teaching aids were also used as material for memorizing and copying practice.

The earliest grammatical texts from the Old Babylonian period were studied and developed in the succeeding periods up to the sixth and fifth centuries BC. The earliest grammatical training texts (from c. 3000 BC, cf. Black 1989:76) consisted of lists of words grouped according to their determinatives. They were later complemented (from c. 1900 BC) by lists containing the signs of the cuneiform writing system. They vary from single-column to four-column lists. They include paradigms, lists of morphemes, word lists, as well as other forms of conveying grammatical information, such as dialogues. The texts were written on clay tablets generally forming large collections known as 'series'.

The analytical method characterizing Old Babylonian grammatical texts is the paradigm. The texts contain bilingual lists (Sumerian on the left and Akkadian on the right) of selected forms arranged according to morphologically determined patterns. As there is less variation in the nominal system, much attention is devoted to the Sumerian verbal system. Sumerian is a highly agglutinative ergative language and listing all potential forms of a verb would be almost impossible. Thus, the structure of the paradigm is often governed by Akkadian, showing the appropriate correspondences in forms of the Akkadian verb.

The earliest examples of paradigms date from c. 2500 BC and they include short lists of contrasting verbal forms inserted in word repertoires used in scribal training (cf. the following examples from Civil 1994:77):

Transitive verb:
in-na-sum 'he will give to'
ì-na-sum 'has been given to him'
nu-ì-na-sum 'has not been given to him'
ḫe-na-sum 'let it be given to'

Intransitive verb:
ba-til 'has been finished'
nu-til 'has not finished'
in-til 'he finishes'
ḫe-til 'let him/her/it finish'

The largest and most complex paradigm is that of the Sumerian verb *gen* (Akkadian *alāku*) 'to go', which has 318 entries.

In addition to the verbal paradigms, there are also long paradigms of pronominal forms with different prepositional and conjunctive elements, etc. It is believed that one of these paradigms must have originally had around 1,200 forms (cf. Civil 1994:77).

Besides paradigms, 'grammatical vocabularies' are the second type of Babylonian grammatical texts. Initially, they were probably simple vocabularies of pronominal, adverbial, and prepositional phrases composed in Middle Babylonian times before the twelfth century BC. Larger systematic lists were developed much later (fifth to fourth centuries BC) by the introduction of paradigmatic groups in which one or more elements were systematically varied. Although their organization is less rigorous than that of paradigms, these lists of morphemes are very helpful for present-day linguists, thanks to their tabular form with translations.

Word lists are the largest number of linguistically interesting Sumerian texts. They date from c. 2800 BC, being found among the oldest tablets discovered in Mesopotamia, and are still found until the end of the third millennium. Originally unilingual, lexical compilations turned Sumero-Akkadian bilingual around the eighteenth to seventeenth centuries BC.

Other types of Babylonian grammatical texts include dialogues used in scribal schools for training purposes, often included in parts of the curriculum. They cover various topics related to writing, spelling, and vocabulary. Some were intended for scribal training, and others for instructing administrators and foremen on how to give orders and conduct businesses in Sumerian.

The grammatical tradition of Babylon offers the oldest instance of linguistic analysis and an invaluable source for the reconstruction of Sumerian. It is interesting to observe with Black (1989:96) that, even if it was a closed system, the Babylonian grammatical tradition spread to other territories as well. It reached the Hittite capital Hattusas and the city-state of Ugarit by the thirteenth or twelfth centuries. It also reached Assyria in the eighteenth or seventh centuries.

References

Black, J.A. 1989. The Babylonian grammatical tradition: the first grammars of Sumerian. Transactions of the Philological Society 87(1). 75–99.

Black, J.A. 1990. Recent study of Babylonian grammar. Journal the Royal Asiatic Society 1. 95–104.

Civil, Miguel. 1994. Sumerian. History of linguistics, Vol. I: The eastern traditions of linguistics, ed. by Giulio Lepschy, 76–87. New York: Longman.

Hymes, Dell (ed.) 1974. Studies in the history of linguistics. Traditions and paradigms, Bloomington, London: Indiana University Press.

LAURA DANILIUC

Balkans

The term 'Balkan' refers to the southeasternmost peninsula of Europe and, collectively, the countries located there. The Balkan Peninsula was named after a mountain range, the Stara Planina in northern Bulgaria whose Turkish name is Balkan (mountain). The relief of the region is shaped by mountain ranges: the Balkan Mountains lie east–west across Bulgaria, the Rhodope Mountains extend along the Greek–Bulgarian border, the Dinaric range extends down the Adriatic coast to the Pindus Mountains in Greece, while in the northwest there are the Julian Alps and the Carpathians. Although there is no sharp separation in the north between the peninsula and Central Europe, the line of the Sava and Danube rivers is commonly considered to be its northern limit. The geographical area covers about 629,000 square kilometers with a population of about 70 million people divided into many countries, including Albania, Bosnia–Herzegovina, Bulgaria,

Macedonia, the mainland of Greece, and the European part of Turkey. Geographically, it also includes parts of Croatia, Slovenia, Serbia and Montenegro (formerly Yugoslavia), and SE Romania, which are often considered Balkan countries as a whole because of their close ties with the region in history and politics.

Linguistic Landscape

In the Balkans, many languages of different genetic affiliation are being spoken belonging to different groups of the Indo-European (IE) family along with some non-IE languages. Most numerous are those of the Slavic family with more than 23 million speakers, including Bulgarian, Macedonian, Bosnian, Croatian, Serbian (the last three were formerly labeled Serbo-Croatian), and Slovenian as national languages, in addition to Slovak, Rusyn, Czech, Polish, Ukrainian, and Russian varieties of the respective ethnic minorities. The languages belonging to the Romance family spoken by about 20 million speakers are (Daco-) Romanian (the official language of modern Romania), Aromanian (also known as Vlach or Macedo-Romanian, the variety of Romanian spoken by minorities in Albania and Greece for at least several centuries), Megleno-Romanian (the variety of Romanian spoken by a minority in Greece), Istro-Romanian, Istro-Romance or Istriot (two archaic languages spoken in Istria on the western part of the Balkans), Italian and Venetian (as minority languages in Croatia and Slovenia), and Judeo-Spanish (in small dispersed communities mainly in Bulgaria, Greece, and Turkey).

Other IE languages spoken on the Balkan are Greek, with about 10 million speakers (in its different dialects and written forms), Albanian, with about 6 million speakers (including its two dialects Tosk Albanian, the official language of Albania, Gheg Albanian spoken in Kosovo as well as Arvanitika—the variety of Tosk Albanian spoken in Greece for some 600 years or more), several idioms of the Indo-Aryan language Romani spoken by about 3 million speakers of dispersed communities in all Balkan countries (Balkan Romani, Vlach Romani, and Sinte Romani), and two Germanic languages—German and Yiddish—spoken by minorities in Romania. There are also non-IE languages that are spoken by minorities on the Balkans—the Uralic, Finno-Ugric language, Hungarian, and a number of Turkic languages belonging to the Altaic family.

This diversity of languages is further reflected in the usage of two alphabets. In the western part of the Balkans, the Slavic languages traditionally have been written in the Latin script (Croatia, Slovenia, and parts of Bosnia and Herzegovina) while the Cyrillic script is used in the eastern parts (Bulgaria, parts of Bosnia and Herzegovina, Montenegro, Macedonia, and Serbia). The use of alphabets is also related to religious influences that have tended to divide the Balkan peoples culturally. The Slovenes and Croats are predominantly Roman Catholic, as are a number of Albanians and Romanians. Islam is adhered to by the Turkish minority, by the majority of Albanians and Gypsies, and by some Bulgarians, while in Bosnia and Herzegovina, and in Macedonia, Montenegro, and Serbia, many indigenous Slavs converted to Islam following the arrival of the Ottoman Turks in the fourteenth century. Elsewhere, the Eastern Orthodox tradition predominates, organized through Bulgarian, Greek, Macedonian, Romanian, Serbian, and Albanian national churches.

Through the long-standing areal contact and extensive exchange of language materials, speakers of typologically distinct languages of the Balkans have adopted features from each other over the centuries, sharing today sets of typological properties termed balkanisms that are not the result of either their common genetic origin or chance. This specific development of the relations between the Balkan languages gave grounds for the term Sprachbund (from German) to be used to denote the particular relations between them in terms of linguistic union. This unity in diversity and similarity without genetic affiliation have remained outstanding features of the linguistic picture of the peninsula, reflecting both the influence of geographical factors and the historical processes of ethnocultural interaction.

Geography

Geographically, the area is composed of two parts: one southern and Mediterranean, and the other northern and continental.

The southern part is open to the easily navigable seas with the Adriatic and Ionian seas on the west, the Sea of Marmora, the Dardanelle and the Aegean Sea on the south, and the Black Sea and the Bosporus on the east. Surrounded on three sides by water, the peninsula has been linked through history with deep and lasting bonds to Anatolia, and Asia beyond, the Italian peninsula to the west, and to the eastern Mediterranean and Egypt to the south. Therefore, it has always been open to various influences and pressures in a permanent exchange of populations to and from all these destinations throughout the whole of its history. This communicative permeability and continuous intimate contacts of its peripheral parts with neighboring non-Balkan areas are contrasted with the central mountainous part characterized by isolation and perpetuation of a most conservative way of life.

Except for the northern plains, which are quite open toward Central Europe, the continental part is

characterized by the mountains that have contributed to the continued fragmentation of human groups in the area. They have been physical obstacles, preventing efforts at regional, economic, and cultural integration dividing the region into small units, in which distinct ethnic groups have tended toward distinct national cultures and languages, local economies, and political autonomy. The mountains have also subdivided the whole area into vertical ecological zones, ranging from lowland farming areas to wooded or rocky uplands acting as places of refuge for ethnic groups expelled from more desirable coast and valley lands throughout history.

Although considerably different, the two parts of the Balkan peninsula are connected by all kinds of historical, cultural, and linguistic links as several rivers crossing the almost continuous mountain barrier provided for continental roads, which, although difficult and dangerous, were always important as lines of communication and channels of exchange of goods and ideas.

History

Since the beginning of its history, the Balkan Peninsula has been very complex from the linguistic point of view, an area of great linguistic diversity in which a variety of different languages have always been in close contact. It has been a crossroads of migration and mixing of peoples for many centuries, so that ethnic diversity has come to be one of its most prominent features. Many different peoples have migrated into the area, absorbing or being absorbed by already existing groups that left traces in the languages spoken there in different layers of language contact.

The whole area shows the same stratification of linguistic elements since early prehistory, which reflects the successive settlement of the region by different populations until today. Archeological and historical data report that the Balkan Peninsula had been an inhabited area even before the immigration of the IE tribes. This pre-IE substratum (or substrata) belonged to an old pre-IE language complex of Mediterranean languages, far related to IE (probably linked with the expansion of agrarian economy along the coast of the Mediterranean around 5000 BC). As they were completely assimilated by later IE expansion, little is known about them, and all opinions about their origin and their relation to the IE family remain purely theoretical. However, there is no doubt that they left traces in all the ancient IE languages of this area. The gradual Indo-Europeanization of the Balkans was a rather long process, which occurred during the second millennium BC. It brought to the region tribal groups speaking several IE languages termed collectively Paleobalkan languages: Illyrian and Venetic to the northwest of the

peninsula, Thracian and Ancient Macedonian to the northeast, Dacian in the north, and Greek to the southeast. With the exception of Greek, very little is known about them from the fragments conserved in ancient Greek manuscripts consisting mainly of toponymic terms, personal names, and some words, aside from the fact that they are all of IE origin. They all seem to be related and have a set of common traits in syntax and morphology with Greek dialects and other languages of the region. The Greeks established colonies outside Greece during the last millennium BC, while during the reign of the Macedonian Alexander the Great in the fourth century BC, much of the Balkan Peninsula came under Greek influence so that the Ancient Macedonian language assimilated entirely by Greek disappeared in the third century BC. This process of Hellenization of the Balkans caused the formation of several common features in all languages that were spoken here later in history.

The Romans first invaded the peninsula in the third century BC, and by the first century AD it was entirely under their control. At the height of Roman power, the Balkan peoples were united under a common legal system, a single ultimate political and military power, and were connected through several important commercial routes. From the third century BC until the time of Slav immigration in the sixth century AD, when the entire region belonged first to the Roman and later to the Byzantine Empire, radical romanization occurred, which is reflected in a number of toponymic traces, while the Latin language of the Roman domination represents a major linguistic substratum of the subsequent Slavic languages.

Venetic was quickly assimilated by Latin under the Roman rule in the first century; Thracian language disappeared in the fifth or sixth century AD, as Thracians were first assimilated by Romans and later by Slavs and Bulgars. The Dacians were assimilated by Romans and are thought to be ancestors to the present Romanians. Some Illyrian tribes were either romanized or assimilated by later Slavic migrations, but others moved south into present-day Albania, where they managed to retain their identity, including the Illyrian language that is said to be the ancestor of the modern Albanian language. Some of the romanized Thracian and Illyrian populations survived by taking refuge in the isolated mountain areas where they lived like shepherds conserving several Balkan Romance varieties.

After the fall of the Roman Empire, this Latin was considerably modified in isolation, developing through absorption of different Paleobalkan substrata in what is today called the Vulgar Latin variants. The mixture of those tongues with later additions of Slavic and other elements created interesting structures and the vocabulary of several Romance varieties still spoken in the

Balkans: Daco-Romanian, Aromanian, Istro-Romanian, and Istro-Romance or Istriot, in addition to Dalmatian, an extinct language spoken along the Adriatic coast until the beginning of the twentieth century.

Later invasions by peoples from the north culminated in the arrival of the South Slavs, who occupied much of the region by the sixth century AD. Consequently, non-Slavic peoples of the Balkans were influenced by Slavic culture, and their languages contain many words of Slavic origin. The Slav invasions were followed by the arrival of other smaller groups. In the seventh century, horsemen called Bulgars settled in the area of present-day Bulgaria, but they became absorbed into the more numerous Slavs and ceased to exist as a separate people. In the ninth century the Hungarians, or Magyars appeared in the Danube Plain, who, unlike most other small groups of invaders, retained their language and culture.

After the fall of the Western Roman Empire in 476, the region of ancient Illyria became part of its eastern portion, the Byzantine Empire, which continued its political, cultural, and linguistic influence through Byzantine Greek on the eastern part of the Balkans for another 1,000 years, until it was defeated by the Turks in 1453. It also introduced Orthodox Christianity to the Romanian and Slavic tribes of this part of the Balkans. For this purpose, in the ninth century, Greek priests Cyril and Methodius translated the Bible from Greek into what is now called Old Church Slavonic, which reflected several Common Slavic features and was at that time widely understood by the Slavs. Translating the Bible, the Byzantine priests introduced plenty of Greek, Latin, and Hebrew words into Slavic.

In the fifteenth century, the Ottoman Empire conquered the peninsula, and a long period of its rule has had a lasting effect on the arrangement and characteristics of the Balkan peoples changing the geographic pattern and cultural characteristics, mainly by forcing migrations throughout the peninsula. For almost 500 years, Albania, Bulgaria, Bosnia and Herzegovina, Montenegro, Serbia, and areas of present-day Romania were under the Turkish control. The area of present-day Croatia and Slovenia in the western part of the peninsula was throughout this period under the political, cultural, and linguistic influence of Italy (Venice) and later German-speaking Austria-Hungary.

The most notable result of these movements of small groups of people and even entire small communities from region to region and small-scale language shifts in certain regions at different times was a radical change in the distribution of various Slavic dialects and linguistic discontinuity in many parts of the Balkans, as well as large-scale multilingualism. During this period, all peoples of the Balkans were strongly influenced by Turkish culture, as reflected in their customs and terminology of dress, food, and music. A large number of Turkish loan words can be referred to as Balkan Turkisms because of their existence in all Balkan languages, Bosnian, Macedonian, Bulgarian, Serbian, Albanian and, to a lesser extent, Romanian and Greek.

The breakup of the Turkish empire in the nineteenth and twentieth centuries led to the establishment of new states that had to find a way of including many nationalities. The recent history of the area reflects a similar fundamental contrast between convergent and divergent tendencies in the development of Balkan cultural and linguistic identity. During the twentieth century, in the multinational state of Yugoslavia including all South Slavs except Bulgarians, Serbo-Croatian as the official language acted as a converging factor, while the later struggle for political independence has contributed to prevailing divergent cultural and linguistic tendencies of various ethnic groups until the present.

Linguistic Union

This common historical fate of the Balkan peoples, their coexistence within the borders of the same empires (Roman, Byzantine, Ottoman), and exposure to similar cultural influences have contributed considerably to the emergence of a shared cultural identity in all fields of human activity including common linguistic features, both lexical and morpho-syntactic. Although the Balkan linguistic phenomena have attracted the attention of numerous linguists since the third decade of the nineteenth century, in spite of their prolonged efforts no general agreement has been reached as to which particular languages belong to the union, the linguistic traits that characterize the area, and the sources of these areal features. It is true, that not all Balkan languages have an equal share in the Balkan Sprachbund and the common properties are most numerous in those parts of the Balkans where the Byzantine and Turkish influence was strong and long term, and where the greatest number of languages are co-territorial. The epicenter of Balkanisms seems to be in the southern bordering area of Macedonia, where Greek, Albanian, Macedonian, Aromanian, and Balkan Romani intersect. Therefore, due to varying distribution of Sprachbund features, many scholars consider five languages including Bulgarian, Macedonian, Greek, Albanian and Romanian, and some southeast Serbian dialects to be the core Balkan languages, while others displaying only some of the features are regarded as less typical and peripheral.

A considerable contribution to the study of these phenomena was offered by Kristian Sandfeld in his classic comparative study of the Balkan languages Linguistique balkanique (1930), in which he recorded

over one hundred Balkan Sprachbund properties, making a distinction between general concordances and concordances between individual Balkan languages. Since then, however, a number of different analyses pointed out different combinations of Balkan Sprachbund properties, although most of them agree along with a number of similarities in vocabulary on one phonological property—the presence of the schwa phoneme, a high or mid central vowel, and six grammatical properties: (1) substitution of case inflections by uninflected function words; (2) grammaticalization of the category of definiteness through definite articles; (3) pronominal doubling of objects; (4) expression of future tense using the verb want; (5) perfect tense with an auxiliary verb corresponding to have; and (6) partial or total loss of the infinitive and its substitution by subjunctive clauses.

In spite of various proposals for source languages of these common features, the exact nature of the contact that led to the Balkan linguistic phenomena remains a matter of some controversy. Several early scholars tried to explain their origin through the substratum hypothesis, suggesting that they developed under the influence of an ancient substrate—that of the Paleobalkan languages Thracian, Dacian, or Illyrian. Other scholars attributed the decisive influence to either Greek or vulgar Balkan Latin. All these hypotheses have been considered to be unrealistic as the Balkan Sprachbund properties developed in the post-Byzantine period and were not present either in classical Greek or Latin, while the Paleobalkan linguistic material is too limited to allow any conclusions of this type. More recent theories suggest multiple causation including the reciprocal influence of the Balkan languages and Slavic influence through massive bi- or multilingualism, sociolinguistic accommodation, or some combination of these factors, which has led to the convergence of different languages.

These contact-induced linguistic changes, however, have not led to a significant, massive restructuring either in syntactic structure or in lexicon, even in the most affected Balkan languages. The Balkan Sprachbund, therefore, has been increasingly seen as a complex of shared features in the respective systems of languages involved, resulting from an interplay of several factors including inheritance, innovation, and contact in a multilingual environment, in which, among the random changes in each language, those were more easily spread that contributed most to direct intertranslatability between them.

The current linguistic tendencies of the Balkans at the beginning of the twenty-first century are closely related to the issues of ethnic identity and are shaped by recent interethnic conflicts resulting in puristic and divergent developmental tendencies in all languages spoken in the region.

References

Banfi, Emmanuel. 1985. Linguistica Balcanica. Bologna: N. Zanichelli.

Fraenkel, Eran, and Christina Kramer (eds.) 1993. Language contact–language conflict, Balkan studies, Vol. 1. New York, Bern, Berlin, Frankfurt/M., Paris, Wien: Peter Lang Publishing.

Joseph, Brian D. 1983. The synchrony and diachrony of the Balkan infinitive: a study in areal, general, and historical linguistics. Cambridge: Cambridge University Press.

Katičić, Radoslav. 1976. Ancient languages of the Balkans (Trends in linguistics). The Hague and Paris: Mouton.

Lindstedt, Jouko. 2000. Linguistic Balkanization: contact-induced change by mutual reinforcement. Languages in contact, ed. by Dicky Gilbers, John Nerbonne, and Jos Schaeken. Amsterdam: Rodopi.

Rivero, Maria-Luisa, and Angela Ralli (eds.) 2001. Comparative syntax of the Balkan languages. Oxford: Oxford University Press.

Sandfeld, Kristian. 1930. Linguistique balkanique. Problemes et resultats. Paris: Honore Champion.

Schaller, Helmut. 1977. Bibliographie zur Balkanphilologie (Bibliography for Balkan linguistics). Heidelberg: Carl Winter.

ANITA SUJOLDZIC

See also **Albanian; Greek; Language: Contact—Overview; Latin; Old Church Slavonic; Romanian; Serbo-Croatian and South Slavic languages; Sprachbund**

Baltic Languages

The Baltic languages, a branch of the Indo-European family of languages, are spoken in the area bordering the Baltic Sea. The principal Baltic languages are Latvian, Lithuanian, spoken on the eastern shores of the Baltic Sea, and Old Prussian (extinct since the seventeenth century). Other extinct Baltic languages are associated

with historical regions of northern Europe, such as Yotvingian, Curonian, Selonian, and Semigallian.

Linguists consider that the Baltic languages are more closely related to Slavic, Germanic, and Indo-Iranian (in that order) than to the other branches of the family. For this reason, some linguists regard Baltic and Slavic as branches of a single Balto-Slavic division of the Indo-European family.

The early common ancestor of the Baltic languages is traditionally referred to as Proto-Baltic and it is believed that it broke off from the other Indo-European languages before 1000 BC. A further division into East Baltic (Latvian and Lithuanian) and West Baltic (Old Prussian) is believed to have taken place before 300 BC.

In terms of phonology and grammar, the Baltic languages are said to be the Indo-European languages closest to Proto-Indo-European. They display a high degree of inflection in both the nominal and verbal systems. Beside Sanskrit, the Baltic languages are sought to be the closest to the origins of the Indo-European languages. The reason is that the Baltic languages, Lithuanian in particular, remained relatively unchanged in time, while other Indo-European languages suffered several alterations. From all Indo-European languages spoken nowadays, Lithuanian has best preserved the ancient sound systems, many morphological and lexical characteristics. Therefore, a basic knowledge of Lithuanian is indispensable in the study of Indo-European languages.

Old Prussian is the most important of the extinct Baltic languages and the most archaic of the recorded Baltic languages. The speakers of Old Prussian were completely assimilated to German territories by the seventeenth century.

The earliest Old Prussian written document is a German–Prussian vocabulary, the so-called Elbing vocabulary, compiled in 1300 AD and consisting of 802 Old Prussian words written in a South Prussian dialect. The most important Old Prussian written records are three catechisms translated from German in the sixteenth century. Every written record in Old Prussian has its own specific orthography.

The Old Prussian nominal system had seven noun types and five cases: nominative, genitive, dative, accusative, and vocative. It did not have the dual number, but only the singular and plural. However, it displayed a neuter gender, lost by Lithuanian and Latvian.

Old Prussian verbs had three separate forms in the plural, but not in the singular. The third person was the same in both the singular and the plural. There were three tenses: present, preterite, and future.

Old Prussian vocabulary was quite similar to Lithuanian and Latvian (closer to Lithuanian than Latvian). The orthography was almost wholly based on the German orthography of that time and was quite inconsistent.

Latvian and Lithuanian are the only Baltic languages still alive. However, Lithuanian is far more archaic than Latvian.

A frequent false impression about the Baltic countries (Estonia, Latvia, and Lithuania) is that they all speak similar languages. In fact, there are three Baltic countries and only two Baltic languages. Latvian and Lithuanian belong to the Baltic subgroup of the Indo-European language family, while Estonian belongs to the Finno-Ugric language family.

Lithuanian

The Lithuanian language, the oldest of the two surviving Baltic languages, is spoken by approximately 3 million people in Lithuania, where it has been the official language since 1918, and by an additional half million elsewhere in the world, in Poland and Belorussia and in such countries as the United States, Canada, and Australia. It is estimated that more than 85% of the population of Lithuania are native Lithuanians.

Lithuanian occupies an important place in linguistics as it is considered to be the most ancient of the living Indo-European languages. It is also the closest language to Proto-Indo-European, the parent language from which all the Indo-European languages emerged. It has inherited many Proto-Indo-European characteristics, such as a pitch-accent system, a rich inflectional case system, and many little-changed Proto-Indo-European word stems. It is believed that Modern Lithuanian is the oldest living Indo-European language; nevertheless, it has one of the most recently established standard literary traditions.

Lithuanian has two main dialect groups, Low (or Western) Lithuanian (along the Baltic coast), with three subdialects, and High (or Eastern) Lithuanian, with four subdialects. Although the country is rather small, it is still rich in dialectal variation. The Low dialect is spoken by the Lowlanders, who live in the west and along the Baltic Sea; High Lithuanian is spoken by the Highlanders, who live in the Eastern (and greater) part of Lithuania.

Even if a literary language had been in use since the sixteenth century, three literary dialects fought for supremacy in the nineteenth century. The most prominent person in the formation of the rules of the standard language was Jonas Jablonskis (1861–1930). Modern Standard Lithuanian is based on the southern branch of the West High Lithuanian dialect and completed its development after the first period of Lithuanian political independence (1918–1940), when it became the country's official language. Lithuanian

experienced a major Russian influence especially after World War II.

Old Lithuanian is known from written documents dating from the sixteenth century, more precisely from 1547 when the first book in Lithuanian was published. Phonetically, Lithuanian went rather far from Indo-European. Baltic languages, as well as Slavic languages, did not preserve aspirated and labiovelar consonants, syllabic sounds, and some vowels including schwa. However, Baltic languages were not influenced by any other non-Indo-European sound system, and basically no new phonemes appeared in Lithuanian.

The most complicated feature of Lithuanian phonetics is stress: there is a system of three tones that is very important in pronunciation. Acute, circumflex, and gravis intonations act for long and short vowels of Lithuanian, and furthermore the stress is absolutely free. The Lithuanian has a free stress in contrast to Latvian fixed stress, which occurs on the first syllable. Lithuanian is rich in the use of diphthongs and, like Latvian, in rising and falling intonations.

Lithuanian is written in the Latin alphabet (33 letters), with additional diacritical marks.

Lithuanian is a highly inflected language. In standard Lithuanian, nouns have seven cases (some dialects have eight or more), three numbers (in classical Lithuanian, while in modern colloquial language the dual number is seldom used), and two genders, masculine and feminine, as neuter disappeared earlier in Baltic languages and was preserved just in some adjective forms. There are twelve inflectional types and five nominal declensions, each of them representing Indo-European nominal stems. The article is not used. Lithuanian adjectives have three declensions.

The Lithuanian pronominal system is a very rich one, with several classes of pronouns: personal, demonstrative, interrogative, attributive, negative, definite, and indefinite. There is also a reflexive pronoun, similar to the one in Romance and Slavic languages.

The Lithuanian verbal system, although less complex than its Proto-Indo-European ancestor, has quite a number of inflected forms. Four simple verb tenses (present, preterite, frequentative past, and future) and several compounds exist in the indicative mood; imperative, subjunctive, reflexive, infinitive, and participial forms are also clearly defined. The frequentative past tense does not exist in Latvian. There are three verbal conjugations. The dual number is also preserved. The third person in Lithuanian, as well as in all other Baltic languages, does not vary in number.

Lithuanian, as well as Latvian, has many compound tense forms, based on the forms of the verb *buti* 'to be' and participles. Lithuanian and Latvian word order is quite free, and, in general, the syntax of both languages is quite similar.

Latvian

Latvian, or Lettish, one of the two surviving Baltic languages, has been the official language of Latvia since 1918, when Latvian independence was won. It is the mother tongue of some 3 million persons living on the eastern shore of the Baltic Sea.

Historically, Latvian has three groups of dialects: Central Latvian, the language of literature; Livonian Latvian, greatly influenced by Livonian and Curonian; and High Latvian (or Latgalian), which underwent considerable Slavic influence. Nowadays, there are two other major dialects: West Latvian (also called Livonian or Tahmian) and East Latvian (High Latvian).

The earliest printed writings in Latvian are translations of religious works: a Roman Catholic catechism, *Catechismus Catholicorum* (1585), a Lutheran version (1586), and a translation of the Bible (1685). In 1638, the first Latvian (–German) dictionary, by Georgius Mancelius, appeared; the first grammar of the Latvian language was published in 1644 by Johann Georg Rehehausen. The works of the Latvians Juris Alunans (1832–1864) and Atis Kronvalds (1837–1875) had a great influence on the development of a standard Latvian language, which was finally established at the end of the nineteenth and the beginning of the twentieth century, based on the Central Latvian dialect.

Latvian is closely related to Lithuanian, but it is less conservative, displaying many sound changes in its historical development. Phonetically, the main difference is that Latvian has its stress always on the first syllable, due to the neighboring Finnish influence. Besides, Latvian has short vowels and monophthongs in the final syllables, while Lithuanian has long vowels and diphthongs. Latvian is more archaic than Lithuanian in the intonations inherited from Proto-Baltic.

Based on the fact that Latvian has undergone a number of linguistic changes that Lithuanian has not, historical linguists generally consider that Latvian is the younger of the two Baltic languages. One of the most important changes is the retraction of main word stress almost always onto the first syllable of a word (an influence from Livonian). Because of the change in the stress pattern of the language, many inflectional endings in Latvian have either shortened or been lost altogether.

Grammatically, Latvian, like Lithuanian, has a rich inflectional morphology, having seven noun cases and six verb declensions. The nominal system lost the dual number and the neuter gender; the instrumental case has the same form as the accusative in the singular and as the dative in the plural. Latvian has six nominal declensions with eight inflectional types.

The verbal system is generally the same as in Lithuanian. It has three conjugations (genetically different). There are three persons, the third of which is the

same (apparently from the time of Proto-Indo-European) in both the singular and the plural (as well as the dual).

The verb in Latvian has three tenses (present, preterite, and future). Distinct characteristics are the descriptive and the obligatory moods. Latvian also has a relatively free word order.

The basic vocabulary is originally Baltic, but there are numerous loan words, mainly from Finnish, German, and Russian. Latvian also had close contacts with Swedish and Polish.

First written in Gothic script, the Latvian language has used the Roman alphabet with diacritical marks to indicate the distinctive quality of some vowels and consonants since 1922. The Latvian alphabet has 33 letters, 11 with diacritical marks. The orthography has undergone numerous changes and improvements, and is now quite closely related to pronunciation.

References

Andersen, Henning. 1996. Reconstructing prehistorical dialects: initial vowels in Slavic and Baltic. Berlin, New York: Mouton de Gruyter.

Barentsen, A.A., B.M. Groen, and R. Sprenger (eds.) 1991. Studies in West Slavic and Baltic linguistics. Amsterdam: Rodopi.

Dambriunas, L., A. Klimas, and W.R. Schmalstieg. 1966. Introduction to modern Lithuanian. Brooklyn, NY: Franciscan Fathers Press.

Fennell, Trevor G., and Henry Gelsen. 1980. A grammar of modern Latvian. The Hague: Mouton.

Hjelmslev, Louis. Études baltiques. 1932. Copenhagen: Levin & Munksgaard.

Magner, T.F., and W.R. Schmalstieg (eds.) 1970. Baltic linguistics. Pennsylvania: Pennsylvania University Press.

LAURA DANILIUC

See also **Indo-European 1: Overview**

Bambara, Mandenkan and the Mande languages

Some 35–40 million people in 15 West African states make use of Mande languages as mother tongues or linguae franca (trade languages). These languages are spoken from Senegal to Nigeria and from Mauritania to the Ivory Coast. The name Mande derives from a phonological variant of the name held by the former kingdom of Mali. The official languages of large kingdoms in West Africa were often Mande, such as Soninke in Gana (seventh to eleventh century), Malinke in Mali (thirteenth to fifteenth century), and Bambara in the kingdom of Segou (seventeenth to nineteenth century).

Going back to Joseph Greenberg's classification, the Mande languages, some 100 in total, have been considered to represent one of the main subgroups of Niger-Congo, the largest African language family. The affiliation of the Mande languages to the Niger-Congo group of languages remains tentative. Unlike other Niger-Congo language groups, Mande shows no signs of noun class or concordance systems. Some authors have demonstrated interesting links to Songhai, a language belonging to the Nilo-Saharan group of languages. They showed similarities in the areas of even basic vocabulary (e.g. a similar word root for 'heart'), of word structure and derivation, as well as in word order. Some authors even identified links to such non-African languages as Basque or Canarian in syntax or vocabulary.

Early classifications subdivided the Mande languages into a Mande-Tan and a Mande-Fu group, based on differences in the word for '10' in the individual languages. The current view basically distinguishes between Western and Eastern languages.

Western Mande is dominated by two-tone languages, i.e. languages that use intonational patterns with two pitch levels to express lexical or grammatical meaning. Languages with three pitch levels are rare and found specifically in the Southwestern group. Mani-Bandama languages (Eastern Mande) usually show very reduced and short word forms. They compensate this loss of information by 3–5 different pitch levels, as seen in the Dan language of the Ivory Coast with 5 levels. While other languages often use the syllable as the tone-bearing unit, Mande languages — and specifically Western Mande — assign tonal patterns to the word. Regardless of the number of syllables, words in the Bambara language will always be limited to one of two tonal structures covering the entire word.

Most Mande languages show a very distinct consonant–vowel (CV) syllable structure, and word stems of 2–3 syllables prevail. Specifically in the Mani-Bandama languages, reduction processes are responsible for many monosyllabic word forms and altered syllable structures. In Gban, numerous words display a ClV syllable structure, such as *klɛ̃* 'Moskito'.

The Western Mande languages show 5–7 oral vowels and, at times, as many nasal vowels (7 oral and 7 nasal vowels in Bambara). Many Mani-Bandama and Volta-Niger (Eastern Mande) languages are particularly complex and make fine distinctions between different vowel sounds. Languages with many different vowels often use vowel harmony, which means that all vowels in a given word must share certain characteristics. For example, the Volta-Niger language Northern Samo of Bangasoko only uses either open (long) or closed (short) vowels within the same word stem. The only difference of words like *zɔ́rɓ* 'remove' and *zùró* 'throw' is vowel aperture. These words could just as simply be transcribed as *zùró^O* and *zùró^C*.

Long vowels in Mande languages are historically most often the result of compensating for the loss of a consonant between two vowels. The Mauka language in the Ivory Coast lost most of its intervocalic consonants, a process resulting in increased vowel length, e.g. sìi 'to sit' (Bambara *sìgi*), sì 'spend the night'. Mani-Bandama developed many very short words, so vowel length became an important distinguishing (phonemic) factor between different words. In dialects of Northern Samo, a Volta-Niger language, there are five different vowel lengths, two of which are phonemic.

Some special consonants occur only in specific geographic areas. Doubly articulated stops such as *gb* or *kp*, or ɓ, ɗ and other implosives, are only found in southern regions, not in the Sahel belt. Dialects of the same language may contain these sounds if they are spread further to the south.

Some Niger-Volta and Mani-Bandama languages seem to lack phonemic nasal consonants altogether. If nasal consonants occur in these languages, they are always followed by nasal vowels, i.e. the nasal characteristic is clearly a quality of the vowel. For example, Northern Samo words (Bangasoko dialect), such as *bèré* 'crust, scab' or *mɛ̀nɛ́* 'snake', differ merely in vowel nasalization, which means that 'snake' should be analyzed as *bèré* and *bɛ̀nɛ́*. In several Mani-Bandama and Volta-Niger languages — e.g. Northern Samo or Dan in the Ivory Coast — nasalization appears to be a characteristic of words as a whole: words are either completely oral or completely nasalized.

In Southwestern Mande languages, the initial consonant of certain word classes changes in specific grammatical constructions. In Mende, the initial consonant of nouns changes in possessive constructions: *ndɛndɛí mia*: 'That is the canoe', but *Mɛndɛ lɛ́ndɛí mia* 'That is the Mende canoe'.

Unlike other Niger-Congo language groups, there are absolutely no signs of noun class systems in Mande. With the exception of a feminine pronoun in Jo (Northwestern Mande), gender is nonexistent in Mande. In comparison to other groups of languages, words undergo very little change due to grammatical influences. Typically, the position of a word in the sentence fully determines its function. Compare *cɛ̀ mùso* 'the man's wife' with *mùso cɛ̀* 'the woman's husband'. In most cases, singular and plural nouns have the same form, provided other logical indications of plurality are present. See Mauka *mɔ́ɔ à shyá* 'The persons are numerou's, literally' '*person is numerous*'.

Nouns and verbs are indistinguishable in form in the predominant majority of the Mande languages. Position alone determines the function of such noun–verbs or neutrals. A noun–verb in a nominal position functions as a noun, and as a verb in a verbal position. Compare Bambara *à bɛ́ kúma* 'He talks' with *kúma dòn* 'It's a speech'. These functions are distinct in a purely tonal respect in several languages. In Northern Samo, each nominalized verb receives a high tone on the last syllable. In Kelinga-Bozo, a Western Mande language, the tonal pattern is inverted in the case of a change from a nominal to a verbal function. Compare ʃyâ, 'laugh' with ʃyǎ 'laughter'.

Western Mande often uses affixes to express meanings equivalent to sentential structures in Eastern Mande. For example, the Northern Samo phrase 'to put somebody into the action of ….', *à ba gònì dà táágó mà* 'he made the man march', literally 'he has put the man in the action of marching' is expressed as *à ye cɛ̀ lá-taama* 'He made the man march.' in Bambara, where the prefix la- expresses causation.

Mande languages have *portemanteau* morphemes that express two grammatical meanings simultaneously. For example, the element *mu* in Susu (Western Mande) in *a mu fa* 'He has not arrived' details the negation of action as well as the perfect tense.

In Guro, Northern Samo, and some other languages, final negated sentence markers do appear in addition to the (already negating) portemanteau morphemes. See Northern Samo *à ǹ dàa kɔ́* 'He has not come.' Both *kɔ́* and *ǹ* (portemanteau morpheme) indicate sentence negation. The final negated sentence markers are therefore highly redundant and easily prone to loss. Borrowings of such sentence markers can frequently be seen, particularly from neighboring Gur languages.

Mande languages place the possessor in front of the possessed. In nearly all these languages, two forms of possession are formally distinguished. The possessor may dominate the relation ('active relation') or be unable to determine the nature of the relation with the possessum ('passive relation'). Compare Bambara *n ká só* 'my house which I bought and which I may sell at any moment' with *n ká bolo* 'my hand (given to me at the moment of birth)'. In some languages, e.g. Northern Samo, there is a three-fold distinction of

associations: mutual social relations (e.g. kinship), passive nonsocial relations and active relations.

All Mande languages have subject–object–verb word order. In most Mande languages, a sentence is interpreted as passive if a normally present direct object is missing. Compare Western Mande Jeli *Lamina-a sibeo kunɛ* 'Lamini eats the meat' with *sibeo-a kunɛ Lamini munu* 'The meat is eaten by Lamini'.

The Mande languages often use nouns such as words for body parts as postpositions, i.e. words with grammatical functions. For example, Western Mande Jeli *kɔŋ* 'belly' can double as the postposition 'inside, in'. Thus, Jeli *bɔgɔ kɔŋ* may both be analyzed as 'in the bush' or 'the belly of the bush'.

Mande languages often use nonverbal structures where other languages may prefer a verbal one: Mani-Bandama Guro *e à da-na* 'He comes' is literally 'He is at the place of coming.'

Several Mande peoples in Liberia and Sierra Leone developed writing systems where each letter stands for a syllable. Only the writing system of the Vai in Liberia attracted broad public acceptance. The N'ko script developed in 1949 by the Guinean Souleymane Kanté has become quite popular in Upper Guinea.

Bambara and Mandenkan (Manding)

The so-called Mandenkan (literally: the language of Mali) or Manding is a denomination for some 40 closely related language variants within Central-Southwestern Mande. Formerly regarded as different languages, these variants are now generally considered to be part of a language continuum. There are no clear language boundaries, although communication between the extremes is rather difficult. Paradoxically, such 'dialects' often show more differences than other clearly distinct 'languages'. A Bambara speaker in Bamako has a much better understanding of Ivory Coast Jula than of the Bambara of Beledougou 30 km away from Bamako. Compare 'first name' in Bambara of Bamako *tɔ́gɔ*, in Dioula of Ivory Coast *tɔ́gɔ*, in Mauka (variant of Malinke) of the Ivory Coast *tɔ́ɔ* and in Bambara of the Beledougou *cwáa*.

Mandekan is subdivided into a western and an eastern branch. Eastern Mandenkan comprises all important lingua francae within the Mande language group: Bambara, Malinke, and Jula. Speakers of one of these variants can communicate with 30–35 millions mother tongue and lingua-franca speakers in West Africa. The majority of the population in Mali, Guinea, Ivory Coast, and Gambia speaks Mandenkan as first or second language; important parts of the population of Senegal, Burkina Faso, Guinea-Bissau use Mandenkan as means of communication. The reasons for the ascent of Mandenkan to the dominant language in that part of Africa can be found in (a) the dynamic Mande traders who, over hundreds of years, spread their language along with their products in multiethnic areas; (b) the enormous prestige of Mandenkan as a successor language of the kingdom of Mali, one of the most important African civilizations; (c) the use of Bambara by the French as a means of communication for their West African troops during World War II; (d) the important role of Mandekan variants in traditional and modern entertainment (griots, music, cinema, etc.); and (e) the accelerated intra-African migration over the past 40 years that has increased the necessity of intraethnic means of communication. More than 2 million people from Burkina Faso have migrated to the Ivory Coast — while only 11% speak Mandenkan as mother tongue in the Ivory Coast, 61% use it as lingua franca.

All Mandenkan variants seem to have two different levels of tone, allowing for up to four meaningful tonal patterns, e.g. Mauka in the Ivory Coast *sá:* 'fontanelle', *sá꞉* 'kind of groundnut', *sà:* 'sheep' and *sà꞉* 'salary, pay'. Odiennekan shows inverted tones compared to the other variants of Mandenkan: *á mà dèn yè* 'He has not seen the child'. (Odiennekan), *à má dén yé* (Bambara). The tones of all known Mandekan variants are based on words rather than syllables.

The sound inventory of language variants depends mainly on the language region. Malinke of Kankan in Guinea shows a doubly articulated stop *gb,* unlike the closely related Malinke of Kita further to the North. 'Heavy' is thus *gbílị* in the Malinke of Kankan, but *gílị* in Kita Malinke. Doubly articulated stops and implosives (ɗ, ɓ) only appear in the southern regions of Mandenkan (e.g. in Mauka). Western Mandenkan has five different oral vowels, while Eastern Mandenkan usually has seven. Western Mandenkan quite often shows older elements, such as intervocalic *t,* which has nearly disappeared in Eastern Mandenkan. Compare 'lion' in Mandinka *jàta* and in Bambara *jàra.*

Mandenkan variants are very much alike in word structure and syntax.

Bambara, which is mainly spoken in Mali and adjacent regions, is the most popular Mande language. Used at different levels in Mali by approx. 90% of the population, it offers excellent communication opportunities in the adjacent Ivory Coast and Burkina Faso, both marked by the lingua franca Jula. Jula is actually a Mandekan word for 'dealer'. Jula therefore denotes not the member of an ethnic group, but rather people who speak Mandenkan traders' languages as mother tongues. As the language of (e.g. cola nut) traders was very often Bambara or Malinke, Bambara and Jula are closely related. The Bambara language is used in all parts of society: in tradition and modernity, in movies like those of one of the leading African movie makers,

Souleymane Cissé. Salif Keita and other worldwide acclaimed musicians prefer to sing in Bambara. The dominant use of Bambara on television, including all aspects of publicity, and in the radio of Mali also decisively contributes to its dissemination.

Today, Bambara (Mandenkan) is one of the most dynamic African languages with rapid expansion tendencies to cover the next few decades.

References

Creissels, Denis. 1983. Éléments de Grammaire de la Langue Mandinka. Grenoble: Université des langues et lettres.

Delafosse, Maurice. 1929. La langue mandingue et ses dialectes (malinké, bambara, dioula), Tome 1. Paris: P. Geuthner.

Dwyer, David. 1989. Mande.. The Niger-Congo languages. A classification and description of Africa's largest language family, ed. by J. Bendor-Samuel, 46–65. New York and London: Lanham.

Ebermann, Erwin. 1986. Kleines Wörterbuch der Bambara-Sprache. Wien: Afro-Pub.

Greenberg, Joseph. 1963. The languages of Africa. International Journal of Linguistics, supplement 29 (1). Part II.

Kastenholz, Raimund. 1996. Sprachgeschichte im West-Mande. Methoden und Rekonstruktionen. Mande languages and linguistics. Köln: Köppe.

Mukarovsky, Hans. 1966 Zur Stellung der Mandesprachen. Anthropos (Vienna) 61. 679–88.

Nicolaï, Robert. 1989. Songhai et mandé. Mandenkan (Paris) 18. 69–80.

Pozdnyakow, K.I. 1980 Yazyki mande. Africana —Afrikansky Etnograifchesky Sbornik 12, ed. by D.A. Olderogge, 173–80. Leningrad: Nauka.

Welmers, William E. 1958 The Mande languages. Linguistic language studies 9th round table meeting, ed. by W. Austin, 9–24. Georgetown University Monograph Series, 11. Washington, DC.

ERWIN EBERMANN

See also **Greenberg, Joseph**

Bar-Hillel, Yehoshua

Yehoshua Bar-Hillel (1915–1975) was an Israeli logician and philosopher of language who made significant contributions in a number of linguistic fields: formal and algebraic linguistics, logical aspects of natural language, and computational linguistics, in particular, machine translation (MT) and information retrieval. His principal essays are included in two collections *Language and Information* and *Aspects of Language*.

In most of his writings, Bar-Hillel's aim was to bridge the 'disastrous' gap between logic and linguistics, believing that linguists (particularly semanticists) had ignored logic to their detriment, and that logicians had ignored linguistics by creating a formal system devoid of any relevance to natural language in actual use. He sought to extend the logical syntax of his mentor Rudolf Carnap, which he believed could serve as 'a methodological and terminological basis for structural linguistics', by describing a formalism for relating lexical items to the specific entities (objects, persons, etc.) they refer to in particular contexts, both in natural and in artificial languages; in this, he anticipated the logical and natural language semantics of Richard Montague. With Carnap he explored the possibilities of a theory of 'semantic information', an extension of inductive and probabilistic logic to the semantics of communication. He also saw logic as a kind of 'universal semantics', providing the framework for expressing relationships among linguistic entities whatever the language (including artificial languages)—in contrast to a linguistics-based 'general semantics'.

Much of his work in the early 1950s was influenced by Norbert Wiener's cybernetics and by the possibilities of the newly invented digital computers to test speculations and theories about logic and language. In May 1951, he was appointed to a position in the MIT Research Laboratory of Electronics, with the task of investigating the application of computers to linguistic work, and specifically MT and information retrieval. He visited all the US groups, wrote the first survey (1951), and convened the first conference (June 1952). The conference was a major factor in launching substantial MT research in the following years. There, Bar-Hillel often expressed 'naïve optimism', which he later regretted; however, he was never a believer in the full automation of translation. For practical reasons, he advocated the collaboration of man and machine, the use of editors to revise MT output, the use of restricted (unambiguous) forms of natural language, and the use of subject-specific dictionaries.

His major contribution to algebraic linguistics was categorial grammar, a 'decision procedure' for identifying constituents in grammatically well-formed sentences, based on the logic of Kazimierz Ajdukiewicz—a further example of bringing logic and linguistics closer. His 1953 essay is recognized as a pioneer article in the field, and he made a number of

important contributions to the theory of formal grammar, much under the influence of Noam Chomsky—a close friend from the 1950s until his death. His theory of categorial grammar was first presented at the MT conference in 1952 as a way of dealing with the syntax of natural language—a topic neglected by most linguists at the time. Later, Bar-Hillel demonstrated that categorial grammar was formally equivalent to context-free grammars, and as such (following Chomsky) inadequate for the description of natural language; he had long recognized its weakness in dealing with discontinuous elements. Consequently (1962), he argued forcefully that MT researchers and computational linguists should base parsers and grammars on Chomsky's type of 'transformational grammar' (i.e. with rules for transforming active sentences into passive, or for showing that *look* and *up* in, e.g. *He looked the word up in the dictionary,* form a single unit 'look up'). However, after his death, renewed interest in nontransformational grammars has shown that categorial grammars can be extended to overcome the deficiencies (cf. Wood 1993), and Bar-Hillel's pioneer work remains influential.

Although a firm supporter of Chomsky's formal grammar theory, he was not uncritical. He disliked the confusing and overworked use of 'theory' for different conceptions and philosophies in linguistics, and he argued for the development of grammars of linguistic performance (language in use) — i.e. parallel to his extension of logic to deal with actual communication. And he was a trenchant critic of the semantic theories of Jerrold Katz and Jerry Fodor, dismissing as simply false their thesis that meaning rules could be exhaustively presented in the form of dictionary entries plus rules of combination, since they had ignored the need for inference rules and for mechanisms to relate words and their external referents.

By the late 1950s, Bar-Hillel's confidence in computational linguistics was waning. He wrote a report for the sponsors of MT research (published 1960), which was highly critical of nearly all current projects. He was particularly critical of those groups adopting statistics-based analyses to 'discover' grammars for computer programs, of those investigating interlinguas (i.e. intermediary language-neutral representations), indeed of any that did not keep to modest and realistic objectives. He was opposed to interlinguas not just in practice (they did not lead to economies of programming) but even more in theory: only a logical semantics could form a sound basis in his view.

His main criticism, however, was directed to the assumption that the goal of MT should be fully automatic high-quality translation, and he included a 'proof' that such a goal was nonfeasible, not just in practice but in principle — arguing that no amount of data (linguistic, encyclopedic, inferential, referential) could ever be sufficient to resolve all ambiguities in texts and to choose the best translations in context. This article was perhaps the single most influential publication in the early history of MT, convincing many outside the field that MT was a misguided activity, and it continues today to be cited as 'evidence' of the impossibility of MT.

Later (1962), he lost his belief that even the man–machine partnership in MT could be cost effective, and the last vestiges of his earlier enthusiasm for cybernetics waned with the realization that the idea of a 'learning machine' was a delusion, since 'all attempts at formalising… inference have completely failed'. On the other hand, later still (1971), he withdrew his harsh judgment about the future of practical MT, conceding that quality in MT could not be an absolute and that, in practice, it varied according to recipients and uses.

Bar-Hillel accepted that he often 'trod on … toes', and he did not, in fact, like the role of devil's advocate, which he often found himself taking. In truth, his criticisms were often expressed forcefully (even bluntly) but they were always pertinent, well-argued, and enlightening.

Biography

Yehoshua Bar-Hillel was born in Vienna on September 8, 1915. He moved to Palestine in 1933, and attended Hebrew University from 1935. He received his M.A. in 'The antinomies of logic' in 1938, and a Ph.D. on 'Theory of syntactic categories' in 1939–1945 (interrupted by four years in Jewish Brigade Group, British Army). He worked as a teacher in high school in 1945–1947, and fought in the Israeli War of Independence in 1947–1949. He was a research fellow at Hebrew University in 1949–1953, and received a visiting fellowship to USA in 1950–1951. Bar-Hillel was appointed research associate in the Research Laboratory of Electronics (RLE), Massachusetts Institute of Technology (MIT) in 1951–1952, and convened the first MT conference (at MIT) in June 1952. He was a senior lecturer in Philosophy at Hebrew University in 1953–1958, and held further appointment at RLE, MIT, 1955–1956. He returned to Hebrew University in 1956 and conducted a study tour of the United States (funded by the US Office of Naval Research) in 1957–1960. He was Associate Professor of Philosophy at Hebrew University in 1958–1961, and visiting professor at the University of California at Berkeley in 1960–1961, and then again at RLE, MIT in 1961. He was Professor of Logic and Philosophy of Science, Hebrew University, in 1961–1975; a member of the Israel Academy of Sciences and Humanities in

1963–1975; secretary and organizer of the third International Congress for Logic, Methodology and Philosophy of Science, Jerusalem, in 1964; visiting professor, University of Michigan, in 1965; visiting professor, University of Southern California, La Jolla, in 1966–1967; president, Division of Logic, Methodology and Philosophy of Science of the International Union of History and Philosophy of Science, in 1966–1968; president, International Union of History and Philosophy of Science, in 1967; visiting professor, University of Konstanz, in 1971; and visiting professor, University of Berlin, in 1972. He died in Jerusalem on September 25, 1975.

References

Bar-Hillel, Yehoshua. 1950. On syntactic categories. Journal of Symbolic Logic 15. 1–16.

———. 1951. The present state of research on mechanical translation. American Documentation 2 (4). 229–37.

———. 1953. A quasi-arithmetical notation for syntactic description. Language 29. 47–58.

———. 1960. The present status of automatic translation of languages. Advances in Computers 1. 91–163.

———. 1966. Four lectures on algebraic linguistics and machine translation. Automatic translation of languages, papers presented at NATO Summer School held in Venice, July 1962, 1–26. Oxford: Pergamon Press.

———. 1964. Language and information: selected essays on their theory and application. Reading, MA: Addison-Wesley.

———. 1970. Aspects of language: essays in philosophy of language, linguistic philosophy, and methodology of linguistics. Jerusalem: Magnes Press, and Amsterdam: North-Holland.

———. 1971. Some reflections on the present outlook for high-quality machine translation. Feasibility study on fully automatic high quality translation, ed. by Winfred P. Lehmann and Rolf Stachowitz, 4pp. Griffiss Air Force Base, NY: Rome Air Development Center, Air Force Systems Command (RADC-TR-71-295).

Bar-Hillel, Yehoshua, J. Malino, and A. Margalit. 1974. On logic and theoretical linguistics. Current trends in linguistics, ed. by T.A.Sebeok, Vol.12: Linguistics and adjacent arts and sciences, part.1: 37–101. The Hague: Mouton.

Carnap, Rudolf. 1934. Logische Syntax der Sprache. Wien: Springer. (The logische syntax of language. London: Routledge, 1937.)

Hutchins, John. 2000. Yehoshua Bar-Hillel: a philosopher's contribution to machine translation. Early years in machine translation: memoirs and biographies of pioneers, ed. by W. John Hutchins, 299–312. Amsterdam: John Benjamins.

Kasher, Asa (ed.) 1976. Language in focus: foundations, methods and systems: essays in memory of Yehoshua Bar-Hillel. Dordrecht: Reidel (Boston Studies in the Philosophy of Science, 43). [Includes substantial bibliography of Bar-Hillel's publications.]

Koppel, Moshe, and Eli Shamir (eds.) 1996. Proceedings [of] the Fourth Bar-Ilan Symposium on Foundations of Artificial Intelligence, June 20–22, 1995, Ramat Gan and Jerusalem, Israel. Menlo Park: American Association for Artificial Intelligence. [Conference dedicated to Bar-Hillel and his ideas.]

Wood, Mary McGee. 1993. Categorial grammars. London: Routledge.

W. JOHN HUTCHINS

See also **Artificial Intelligence; Chomsky, Noam; Computational Linguistics; Machine Translation**

Basque

Basque is the oldest continuously spoken language in western Europe. It was already an ancient language, possibly a remnant from the Stone Age, when the Romans first identified it over two millennia ago. Crucial to its survival has been the historically remote habitat of its speakers in the rocky fastness of the western Pyrenees Mountains between Spain and France. The Basque language is related to no language family now extant. Its ancient, remote, and singular nature has intrigued generations of scholars, producing as much speculation as fact.

In Basque itself, the word for the language is *euskara*. Speakers of *euskara*, according to the most recent census data, number over a million people. However, only about two thirds use it as their daily primary language. North central Spain is the home of the overwhelming majority of these speakers. Along a bordering area in southwest France is a smaller group of about a hundred thousand speakers.

The term Basque describes not only a language but also a distinct people and cultural region. This region is called 'Basque Country' (*Euskal Herria*); however, it is not a sovereign country. The effort of some to establish such an independent state has dominated both recent and past Basque history. The struggle has been as futile as it has been violent.

The Basque Country has traditionally lain along the French and Spanish sides of the farthest western edge of the Pyrenees, descending to the Bay of Biscay. The size of the area occupied by the Basques has grown

smaller over the millennia. Basque Country today extends east from around the area of the Spanish city of Bilbao, past San Sebastian, and then veers north at the terminus of the Pyrenees. In France, it extends up past Biarritz, ending around Bayonne. In length, therefore, Basque Country is no more than a hundred miles, with its extent inland being only several dozen miles.

Its total area is a narrow, curvilinear band of less than 4,000 square miles, smaller than the state of Connecticut. In antiquity Basque speakers extended much farther north into France and farther east into the Pyrenees. The Romans identified them as '*vascones*'. The Spanish termed them '*vascos*', and the French, hearing the 'v' as a 'b', as '*basques*'. With the influence of Norman French in England, they were termed Basques in English.

Spain has the largest number of Basque speakers. The urban industrial complex of Bilbao is the largest Basque city, and is the fourth largest city in Spain. Spanish Basque Country comprises three north central provinces, Bizkaiko, Gipuzkoako, and Araba. While not independent, this Spanish Basque Country holds the status of an Autonomous Region, under the constitution of Spain. It is allowed limited local governing authority and elects its own 'president', still recognizing nonetheless the central authority of the Spanish monarchy. *Euskadi* is the name for the Basque political state. There are also some Basque speakers in the mountainous upper parts of the eastern province of Navarra, and in the wine-growing area southward of Riojo. In France, Basque speakers are only in the far southwest province of Pyrénées-Atlantiques and have no autonomous status. Essentially, therefore, Basque Country consists of four provinces in Spain and one in France. The total population is three million people, but much less than a third of these use Basque for daily communication. *Euskal Herria* describes not only the cultural but also the geographic extent of the Basques.

Basques emigrated to the New World in the last century. Leaving their mountain pastures as noted shepherds, a considerable number settled in Nevada. This state holds one of the major communities of Basques outside Spain.

In traveling through Basque Country, one is more likely to hear the language in the rural interior rather than the urban, coastal area. The latter has given Basques a modern prosperity through the steel industry of Bilbao and tourist beach resorts. However, it is in the farms and pasturelands of the interior that Basque has maintained its long-term economic and social survival. Centuries-old stone residences stand as proud homesteads (*baserri* and *borda*) that have been inhabited by the same families for many generations. Here lie the vigor and longevity of Basque.

Surviving from before the invasion of Europe by the Indo-European family of languages that dominate the continent today, Basque falls on the ear as a somewhat hard, guttural language. In looking at a Basque text, one is struck by how often the letter 'k' appears.

Its morphology, in the marking of nouns and verbs for person, number, or time, occurs predominantly by adding a suffix, rarely a prefix. Gender is not normally distinguished. Verbs are primarily transitive, although there are also intransitive verbs and some that are both. The language includes adjectives and pronouns. Moreover, in terms of number there is a distinction between singular and plural, and there is agreement between subject and verb.

Basque syntax uses a standard sentence pattern of subject in the first position, an object in succession, and a verb following: S–O–V. The sentence 'Ann hit John' becomes 'Ana Jonek jo zuen.' Clauses also end in verbs. There are, however, no relative pronouns.

The Basque vocabulary preserves the most ancient pre-Indo-European origin of the language. There are an exceptional number of words for or related to stone. Nevertheless, it has successively accumulated words over the millenia from, for example, Celtic, Latin, and Spanish. The word for 'hello' in Basque is 'augur', derived from the Latin, 'ave'. The initial nucleus of the language was probably in southwestern France, in the province of Acquataine. With the Roman invasions of Hispania and Gaul, Basques retreated south, moving higher and more deeply into the safety of Pyrenees, the strategy whereby the language survived.

There are two major centers for the study and research on Basque culture, history, and language. The first is the University of the Basque Country (in Spanish: *Universidad del Pais Vasco*; in Basque: *EuskalHerriko Unibertsitatea*), with branches in the three Spanish Basque provinces. The other is the Center for Basque Studies at the University of Nevada in Reno.

References

Hualde, José Ignacio, Joseba A. Lakarra, and R. L. Trask (eds.) 1995. Towards [sic]a history of the Basque language. Amsterdam: John Benjamins Publishing.

King, Alan. 1994. The Basque language. Reno: University of Nevada.

Kind, Alan. and Begotxu Olaisola Elordi. 1996. Colloquial Basque. London: Routledge. (with audio cassettes).

Michelena, Luis [Koldo Mitxelena]. See any of the many dozens of works by him, the most noted authority of all time on Basque.

Saltarelli, Mario. 1989. Basque. Croom helm descriptive grammar series. London: Routledge.

Tovar, Antonio. 1957. The Basque language, translated by Herbert Pierrepont Houghton. Philadelphia: University of Pennsylvania Press.

Trask, R.L. 1997. The history of Basque. London: Routledge.

EDWARD A. RIEDINGER

Baudouin de Courtenay, Jan Ignacy Niecisław

Jan Ignacy Niecisław Baudouin de Courtenay was a pioneer of modern linguistics. He anticipated most of the basic concepts of structuralist theories. In particular, his formative influence on modern phonology is generally acknowledged. However, his contribution to linguistic science is sometimes undervalued since he was not able to fully exploit his own discoveries and never published a major synthetic work.

Baudouin de Courtenay is usually credited with having introduced the linguistic term 'phoneme' to denote a speech sound that distinguishes meaning. He initiated the discussion of what makes sounds distinctive in 1870, in his inaugural lecture at the University of St. Petersburg. In his hands, phonetics became more abstract. He pointed out that the sounds of language are structural entities, rather than mere phonatory and auditory phenomena. Baudouin de Courtenay attempted to relate the study of phonemes to the linguistic intuition of speakers. He noted that the way we perceive different sounds is not always a matter of their physical and physiological properties. The motor-acoustic properties are important elements of linguistic structure as long as they contribute to the function of sounds.

Around the turn of the century, international linguistic thought was strongly influenced by the neogrammarian school, centered on Leipzig. Many of Baudouin de Courtenay's ideas were inspired by opposition to the nineteenth-century philology, concerned exclusively with the history of languages. In particular, he stressed the importance of synchronic analysis and suggested a distinction between language and speech (similar to Ferdinand de Saussure's *langue/parole* dichotomy). Still, Baudouin de Courtenay's views of psychology remained within the dominant framework. He considered association psychology a fundamental achievement of modern science. Therefore, Baudouin de Courtenay tried to build a bridge between his functional definition of the phoneme and psychologism. It resulted in a conception of the phoneme as the psychological equivalent of a speech sound. This mentalistic approach to phonology might be considered parallel to the basic ideas put forward by generativism many years later.

Baudouin de Courtenay proposed a new linguistic discipline, 'psychophonetic phonetics' (corresponding to modern phonology). It would have to occupy itself with the relation between sound and the psychological aspects of audition (as opposed to 'anthropophonic phonetics', studying the production and the audition of sounds). In 1895, Baudouin de Courtenay published his chief work, *Versuch einer Theorie phonetischer Alternationen* (1895; *Essay on a Theory of Phonetic Alternation*). Despite its occasional ambiguity, the views expressed in this book have become part of phonology as we know it today. At the beginning of the twentieth century, the idea of the phoneme started to infiltrate into international linguistics. It was later advanced by such linguists as Ferdinand de Saussure, Nikolai Sergeevič Trubetzkoy, Leonard Bloomfield, and Edward Sapir. Taking the lead from Baudouin de Courtenay, they cleared his theory of its psychological overtones. The distinction between the phoneme and the sound proposed by the Polish linguist may be said to have laid the foundations of structural linguistics.

It is worth noting that Baudouin de Courtenay's interest in what we might call psycholinguistics resulted in many analyses of his own children's speech (over 12,000 pages of manuscripts—most of them still unpublished).

Baudouin de Courtenay advocated a bold and individual theory of the nature of language mixing and change. It was put forward in his article 'O smešannom xaraktere vsex jazykov' (1901; On the Mixed Character of All Languages) and might be said to be a reaction to Schleicher's tree image of Indo-European languages. Baudouin de Courtenay rejected the idea of a pure language and attempted to analyze the mechanism of linguistic convergence. He claimed that languages mix both in space and in time (all languages mix with their earlier stages). When languages of different types mix, they can influence each other. Under such influence, a language may undergo many grammatical changes—it may lose inflection, level paradigms, regularize stress, etc. Therefore, in Baudouin de Courtenay's opinion, it is language contact that causes language change.

Although he occupied himself mainly with general problems—questions of theoretical phonetics, language mixture, relations between linguistics and psychology, etc.—Baudouin de Courtenay was also a specialist in comparative and historical linguistics. He pioneered the issues of Old Polish in his Ph.D. dissertation. Since then, the history of the Polish language was among his permanent interests. He was also an outstanding Slavist. One of the three major palatalizations of velar

consonants that took place in Proto-Slavic is named after Baudouin de Courtenay since it was first described by him.

Baudouin de Courtenay was a very charismatic teacher. He exerted a huge influence on several generations of students. His lectures were usually very informal and thought-provoking. They represented the most advanced linguistics to be found in eastern Europe at that time. Baudouin de Courtenay's long teaching career included professorships at several universities of eastern Europe. It began in 1874 when he was appointed to the faculty of the University of Kazan and lasted until his death. Many of de Courtenay's students became prominent linguists. A number of them were active in OPOJAZ (the most important Russian circle devoted to literary theory and modern linguistics), e.g. Evgenij Dmitrevič Polivanov (1891–1938), one of the pioneers in historical phonology. Also, Kazimierz Nitsch (1874–1958), an outstanding Polish dialectologist, was among the pupils of Baudouin de Courtenay. Most importantly, Baudouin de Courtenay's theories were strongly influenced by his favorite student, Mikołaj Kruszewski (1851–1887). It is very difficult to separate the ideas of the two linguists. Therefore, they are often referred to as the theories of 'the Kazan School'. The time spent in Kazan was the most prolific period in Baudouin de Courtenay's life. However, partly due to Kruszewski's untimely death, Baudouin de Courtenay did not have any direct successors.

Baudouin de Courtenay's absorbing interest in linguistics did not deter him from becoming a political activist. He supported emancipatory and democratic movements. After 1918, when Poland regained independence, he emphasized the importance of securing the civil rights of national minorities (the minorities nominated him as their candidate for the presidency of Poland). It was also reflected in his linguistic research—he collected data on many minority languages (e.g. Kashubian dialects in Poland, Slovene dialects in Italy, local dialects in Slovakia, and Yiddish in eastern Europe).

Biography

Jan Ignacy Niecislaw Baudouin de Courtenay was born in Radzymin, near Warsaw, Poland (then a province of the Russian Empire) on March 13, 1845. He obtained his M.A. (1866), Warsaw (at the Department of Philology and History). He received a scholarship of the Ministry of Education for outstanding achievements and continued his studies in Prague (Bohemia), Jena, and Berlin (Germany), tutored by August Schleicher; he obtained his Ph.D. for work on analogy in Polish declination, University of Leipzig, Germany,

1870, and then worked in St. Petersburg, Russia, on Old Polish and Indo-European, in 1870–1874. He was Professor of Comparative Grammar of the Indo-European languages, University of Kazan, Russia, in 1875-1883 (there tutored Mikołaj Kruszewski). He moved to Dorpat (now Tartu, Estonia), where he was Professor of Comparative Grammar of the Slavic languages, in 1883–1893; he was also Professor of Comparative Linguistics and Sanskrit, Jagiellonian University, Cracow (then the capital of the Austrian partition of Poland), in 1893–1899. He lost his teaching appointment because of his radical articles concerning social policy in the Austrian partition of Poland. He moved again to St. Petersburg, where he worked as Professor, University of St. Petersburg, in 1900–1918. He was imprisoned for three months for publishing a pamphlet on autonomy of nations in 1914. He returned to Poland after World WarI in 1918. and worked as Professor of Comparative Linguistics, Warsaw University, in 1920–1929. He stood as a candidate in a presidential election in 1922. He died in Warsaw on November 3, 1929.

References

Adamska-Sałaciak, Arleta. 1996. Language change in the works of Kruszewski, Baudouin de Courtenay and Rozwadowski. Poznan: Motivex.

Adamska-Sałaciak, Arleta. 1998. Jan Baudouin de Courtenay's contribution to general linguistics. Historiographia Linguistica XXV.

Baudouin de Courtenay, Jan Ignacy Niecislaw. 1868. Einige Fälle der Wirkung der Analogie in der polnischen Deklination. Beiträge zur vergleichenden sprachforschung auf dem gebiete der arischen, celtischen und slavischen sprachen VI.

———.1870. O drevnepol'skom jazyke do XIV stoletija.

———.1877. Otčěty o zanjatijax po jazykovedeniju v tečenie 1872 i 1873 gg.

———.1895. Versuch einer Theorie phonetischer Alternationen. Ein Kapitel aus der Psychophonetik.

———.1901. O smešannom xaraktere vsex jazykov. Žurnal Ministerstva narodnego prosveščenija 337.

———.1910. Les lois phonétiques. Rocznik slawistyczny 3.

———.1922. Zarys historii jazyka polskiego.

———.1963. Izbrannye trudy po obščemu jazykoznaniju, 2 vols.

———.1972. A Baudouin de Courtenay anthology. The beginnings of structural linguistics.

———.1974–1983. Dzieła wybrane, 6 vols. Warsaw: PWN.

Jakobson, Roman. 1960. Kazańska szkoła polskiej lingwistyki i jej miejsce wświatowym rozwoju fonologii. Biuletyn Polskiego Towarzystwa Językoznawczego 19.

Jakobson, Roman. 1929. Jan Baudouin de Courtenay. Slavische Rundschau 1. 809-12, reprinted in Portraits of linguists: a biographical source book for the history of western linguistics, 1746–963, Vol.1: from Sir William Jones to Karl Brugmann, ed. by Thomas A. Seboek. Bloomington: Indiana University Press, 1966.

Jakobson, Roman. On language, ed. by Linda R. Waugh and Monique Monville-Burston. London and Cambridge, MA: Harvard University Press.

Rothstein, Robert A. 1975. The linguist as a dissenter: Jan Baudouin de Courtenay. For Wiktor Weintraub on the occasion of his 65th birthday, ed. by Viktor Erlich et al. The Hague: Mouton.

Stankiewicz, Edward (ed.) 1972. Baudouin de Courtenay: his life and work, introduction in a Baudouin de Courtenay anthology. The beginnings of structural linguistics. ed. by Bloomington: Indiana University Press.

Stankiewicz, Edward. 1976. Baudouin de Courtenay and the foundations of structural linguistics. Lisse: Peter de Ridder Press.

Toman, Jindřich. 1991. Nationality as choice—Baudouin de Courtenay's individualistic approach. Cross Currents 10.

PAWEŁ RUTKOWSKI

See also **Bloomfiled, Leonard; Phoneme; Phonology**

Belgium

Belgium is a small parliamentary and constitutional monarchy in northwest Europe; it shares borders with the Netherlands (north and northeast), France (south and southwest), Germany (east), and Luxembourg (southeast), with the North Sea forming a natural boundary on the northwest. It stretches over 32,547 square kilometers and is divided into ten provinces: five French-speaking provinces in the south (Walloon Brabant, Hainaut, Namur, Liège, and Luxembourg), and five Flemish-speaking provinces in the north (Antwerpen, Flemish Brabant, Limburg, Oost-Vlaanderen, and West-Vlaanderen). The Belgian population numbers 10,213,752 (1999). From a linguistic perspective, Belgium has a dual identity, straddling the divide between Romance and Germanic languages. Three national languages are officially recognized: Dutch, French, and German.

Belgium became independent from the Netherlands in 1830 and it was accepted internationally as a neutral state on the 20th of January the following year. French became the language of the government, although Germanic vernaculars (Flemish dialects) were spoken in the north and east of the fledgling state. (German dialectal varieties were used in what is known as 'Alt-Belgien', an area north of Liège around Vielsalm and Arlon, along the Luxembourg border.) At the time, the higher bourgeoisie and the nobility throughout Europe used French as the language of culture. Moreover, part of the population spoke Romance vernaculars (Walloon, Picardy dialect, and Gaumais) and no other major language appeared to be a reasonable alternative: the Flemish dialects showed no unity, and the closest language, Dutch, was rejected as the Belgian Revolution broke out against Holland.

In the 1830s, the prominent position of the French language in Belgium was hardly challenged as linguistic and political factions were all united in a desire to firmly establish the new country as an internationally recognized power, despite attempts made by William of Orange, the Dutch king, to re-annex Belgium. However, this exemplary Belgian unity did not last very long: Flemish intellectuals started to voice claims for cultural freedom in the late 1830s. They were motivated mainly by a desire to protect the Flemish cultural heritage. The movement took on a more social slant in later years. In the 1870s and 1880s, three important linguistic laws were promulgated. First, in 1873, Flemish citizens were allowed to stand trial in their mother tongue if they wished. This came as a result of unfair trials in the 1860s where the accused were condemned in French, a language they did not understand. Then the administration in Flanders became bilingual (1878), which disqualified Francophone civil servants from taking positions in the north of the country. Finally, part of the teaching in secondary schools of Flanders (1883) was to be undertaken in Flemish. With the growing influence of Flemings in the government of the country, a Walloon movement started to emerge, born of the worries of the Romance population who feared the loss of their privileged position within the nation.

In 1898, Flemish claims were crowned by the so-called *loi d'égalité* (the Equality Law) that established the equivalent of the Flemish and French versions of laws. The weight of Flanders in the national political life was further increased when universal suffrage was granted to all men after World War I. Given that the Flemish population was more numerous, Flemings started to occupy strategic functions in the government. Also, the linguistic composition of the country was further complicated by annexation of new German-speaking territories as a result of the Treaty of Versailles (June 28, 1919). This territory, 'Neubelgien', comprises a southern and a northern part separated by the Walloon cantons of Eupen, Malmédy, and St. Vith. The Flemish- and French-speaking populations steadily grew apart during the twentieth century. In 1912, the 'Lettre au roi' (Letter to

the King) by Jules Destrée, a Walloon socialist, illustrates the growing estrangement between the two parts of the Belgian population: it states that there are no Belgians, but only Flemings and Walloons. In 1932, a new law drew a linguistic border between the two main linguistic regions. This boundary could still be modified according to the results of the census. However, the linguistic questions of the census were suppressed after World War II following their boycott by Flemish mayors. Flemings were concerned about losing Brussels—historically a Flemish city but with an increasing Francophone population—and seeing the Francophone 'oil slick' further expanding in surrounding areas. The Gilson laws painfully established the definitive linguistic border in 1962–1963: they included facilities (the possibility of getting state services in the other language in regions where there was a linguistic minority of at least 30%) for Francophones in Brussels (Drogenbos, Kraainem, Linkebeek, Rhode-St.-Genèse, Wemmel, Wezembeek-Oppem), for Francophones in Flanders (Messines, Espierres, Helchin, Renaix, Biévène, Herstappe, and the well-known Fourons), Flemish in Wallonia (Comines, Dottignies, Enghien, Flobecq, Herseaux, Marcq, Petit-Enghien, and Warneton), and German in Wallonia (the six 'Malmedian communes': Belleveaux-Ligneville, Bevercé, Faymonville, Malmédy, Robertville, and Waimes). The cohesion of the country was thus weakening and it was necessary to reorganize the state. In 1993, Belgium became a federal state after a series of reforms: creation of communities (cultural attributions) in 1970, reinforcement of communities and creation of regions (territorial responsibilities) in 1980, and creation of a third bilingual region, the Brussels Metropolitan Region (Bruxelles Capitale), in 1988–1989.

Present-day Belgium is therefore a federal state that consists of three regions— Flanders, Wallonia, and Brussels— and three communities—Flemish, French, and German. Regions and communities do not correspond, which greatly complicates politics in Belgium. From a linguistic point of view, there are four areas: Flanders speaks Flemish only, Wallonia is Francophone—except for the German region—and Brussels is bilingual, although now mainly Francophone and very cosmopolitan, given the presence of bodies such as NATO and the European Union. Lately, the *Communes à facilités* have been a bone of contention. The Peeters' Circular (1997) tried to suppress facilities for French speakers around Brussels: Flemings do see facilities as a temporary solution to allow integration in the other community, while Francophones understand them as a permanent state that could possibly be extended to other regions (especially around Brussels). This has led to ongoing tensions

between the two parts of the country, and in May 1998, the European Commission had to send a Swiss representative, Domino Columberg, to evaluate the situation in Brussels and write a report on the whole linguistic conflict in Belgium and its possible solution.

By the decree of December 10, 1973, Dutch is the official language of the Flemish Community (5,912,000 inhabitants in Flanders and between 13 [Francophone sources] and 25% [Neerlandophone sources] of Brussels' 953,000 inhabitants). Although a Language Union Treaty was signed between Flanders, Belgium, and the Netherlands on September 9, 1980, Flemish Dutch and Netherlandic Dutch are not identical. The main differences are found in pronunciation, but this does not hamper communication. Lexical differences also appear between the two variants, and Flemish Dutch possesses a more Frenchified vocabulary. There are also minor differences in sentence structure. Flemish dialects are still very vivid, although Dutch is the cultural vector. Standard French in Belgium is nowadays the mother tongue of Wallonia (3,327,000 inhabitants) and of most Belgians and immigrants in Brussels. Romance vernaculars are still used in informal situations among older people and working classes; they can also be used by educated people for sentimental, cultural, or amusement purposes. In any case, they are not the exclusive communication means of any Belgian of the twenty-first century.

French in Belgium presents a series of features that distinguish it from Standard French (which is identified as the Parisian variety). Belgian French—as well as regional varieties in France—possesses specific words influenced by the dialects spoken in the area. However, French in Belgium presents more specific features that are linked with the institutions of the country. This is the case, for example, in *bourgmestre* 'mayor' (*maire* in France) or *candidatures* 'first two years of a degree' (*DEUG* in France). In addition to these, Belgian varieties of French contain more words with foreign origins: Germanic in words such as *kermesse* 'kind of seasonal fairground' (*fête foraine* in France) or *couque* 'bread roll' (*petit pain* in France), Spanish as, for example, in *bodega* (bar found in a cellar) or *escavèche* (a fish dish). Because of the numerous powers that have held sway over the territory of contemporary Belgium, French in Belgium also seems more open to English than the Parisian variety.

Popular varieties of French in Belgium have features of sentence structure that seem to be influenced by Flemish: the position of the adjective before the noun (*une propre chemise* 'one's own shirt' instead of *une chemise propre*); interrogation in *qu'est -ce que c'est pour* (based on the Flemish *wat is het voor* 'what is that for'), or replacement of *pouvoir* 'to be able to'

by *savoir* 'to know'. Finally, some words exist both in France and Belgium but have different meanings in the two countries. For example, *torchon* refers in France to a tea towel but in Belgium to a cloth used for cleaning the floor; *un rideau* is a curtain in France and a net in Belgium.

In spite of some differences, Belgian and French citizens understand each other without problems. Nowadays, specificities tend to fade away because of the development of mass media, but there has always been a great interest in language correction in Francophone Belgium. A feeling of linguistic insecurity (the conviction of not speaking the valued variety of a language) has resulted in the development of a strong grammatical tradition in Belgium, of which Maurice Grevisse's grammar, *Le bon Usage*, is the best-known example. There was a general tendency to imitate French usage until recently; but now, Francophone Belgium appears more autonomous, as shown in the law on the feminization of job descriptions passed in 1993, without waiting for French legislation.

German is spoken only by about 1% of the population. A separate German variety has not developed in Belgium, although there are some specificities. These come first from national institutions; for example, *Permanent-deputation* is used instead of *Länder-regierung*. The differences from Standard German also come from interferences with other vernaculars: Germanic dialect, Walloon, and especially French. These 'gallicisms' appear in borrowed words (e.g. *renovieren*, based on French *rénover* instead of German *erneuern*) and in literal transpositions of phrases (e.g. the French *téléphoner à* has changed the Standard German *telephonieren mit* 'telephone with' into *telephonieren an* 'telephone to').

Languages have thus played a key role in the short history of Belgium. Linguistic tensions have shaped the institutional evolution of the country in making a national state into a federal state. Although the country is nowadays divided between the two main linguistic groups with increasing autonomy, linguistic questions (mainly concerning the future of Brussels) still constitute a major driving force in Belgian politics and are likely to remain so in the twenty-first century.

References

Baetens-Baerdsmore, Hugo, and E. Witte (eds.) 1986. The interdisciplinary study of urban bilingualism in Brussels. Clevedon/Philadelphia: Multilingual Matters.

Bal, Willy, et al. 1994. Belgicismes. Inventaire des particularités lexicales du français de Belgique. Louvain-la-Neuve: Duculot.

Blampain, Daniel, et al. 1997. Le français en Belgique. Louvain: Duculot.

Labeau, Emmanuelle (ed.) 2000. France–Belgique: des frères ennemis de la langue de chez nous?, Québec: Centre international de Recherche en Aménagement linguistique (CIRAL).

Lemaire, Jacques (ed.) 1989. Le français et les Belges. Brussels: Éditions de l'Université de Bruxelles.

Moreau, Marie-Louise, et al. 1999. Les Belges et la norme. Analyse d'un complexe linguistique. Brussels: Ministère de la culture, Service de la Langue française.

Pohl, Jacques. 1979. Les variétés régionales du français: études belges (1945–1977). Bruxelles: Éditions de l'Université de Bruxelles.

Verdoodt, Albert. 1977. Les problèmes des groupes linguistiques en Belgique. Louvain: Peeters.

Wright, Sue (ed.) 1994. Languages in contact and conflict: experiences in the Netherlands and Belgium. Clevedon: Multilingual Matters, Current Issues in Language and Society, Vol.1, No.2.

http://cfwb.be – site of the French Community of Belgium.

EMMANUELLE LABEAU

See also **French Language**

Benveniste, Emile

Benveniste is best remembered today for his writings on general linguistics and on semiology (following Saussure's usage, this term has been used in Europe more than 'semiotics'). However, the bulk of his publications, of which we can only mention the most important, focus on comparative Indo-European philology. After the death at the front of fellow-student Gauthiot, Benveniste completed his grammar of Sogdian (published in 1929) and then Meillet's grammar of Old Persian (1931). In his thesis *Origines de la formation des noms en indo-européen* (1935), which deals with Indo-European root morphemes, Benveniste broadens out from Iranian languages to the wider Indo-European field. Notes for the sequel to this research, dealing with suffixal modals, were lost when Benveniste's apartment was ransacked during World War II, but were painstakingly reconstructed and published in 1948. Probably the best known of his works

in this area, both because of its less technical nature and because of the fascinating link between language and culture, is his *Le vocabulaire des institutions indo-européennes* (1969; English translation 1973), in which six sections deal with topics such as the vocabulary of kinship or of religion.

In his work on general linguistics, Benveniste elucidated and further developed the foundations laid by Saussure. He is a proponent of the structuralist notion of a system founded on a series of differences; he promotes the new discipline of semiology, and he extends the remit of French linguistics to cover *discours* (discourse) and *parole* (speech). Benveniste participated in the interwar debate on the arbitrary nature of the linguistic sign, maintaining that the link between the sign and external reality is arbitrary, although the link between signified and signifier is a necessary one. Many of Benveniste's most important articles are collected in his *Problèmes de linguistique générale I & II* (1966, 1974; English translation of Vol. 1, 1971). The articles of the first volume were selected by the author himself, but the selection for the second volume, compiled with Benveniste's help when he was in very poor health, has sometimes been questioned. It is particularly for the pioneering work on *énonciation* that the *Problems in General Linguistics* is read today. The French tradition in *énonciation,* aspects of which are dealt with under pragmatics and speech-act theory in English-speaking circles, treats language not as a static product, but as an interlocutionary and intersubjective phenomenon. Thus, Benveniste, in analyzing the personal pronoun system, points to the difference between the I–you pronouns, which are discoursal units, and the third-person pronouns. The first- and second-person pronouns, moreover, are linked to a deictic system of demonstratives, and spatial and temporal adverbs, as well as to the verbal system. French tenses are analyzed by Benveniste according to a distinction between *histoire* (story/history) and *discours* (discourse), adverbs such as 'here' and 'today', for example, relating to a first-person pronoun and being anchored in the discoursal tense system (whereas in a narrative we would hear that X was in such a place on such a date). This work on pronouns and on tense has been influential in both linguistics and literary criticism. In the former, it feeds into various French traditions, from the linguistics of Antoine Culioli to French discourse analysis. The *histoire/discours* distinction was used to great effect (as *récit/discours*) by literary critic Genette among others. It has been pointed out that Benveniste's work should not be seen as fragmented, because even in his work on historical linguistics he shows the same concern to link language and society and to account for the subjective nature of language, as in his writings on general linguistics.

Biography

Benveniste was born in 1902 in Aleppo, but when he was a baby the family moved to Paris. He attended French and rabbinical schools and then went on to study Indo-European philology, and classical, Indian, and Celtic languages at the Sorbonne. Like his master, Meillet, and Saussure before that, he taught at the Ecole Pratique des Hautes Etudes, taking up a post in Comparative Grammar and Iranian in 1927. Ten years later, he succeeded Meillet in the chair of Comparative Grammar at the Collège de France, having defended his thesis on the origin of Indo-European noun formation in 1935. During World War II, he put his scholarly knowledge to practical use, fleeing to Switzerland with the aid of a priest who communicated with him in Sogdian. He became secretary of the Société de Linguistique de Paris, and editor of the society's bulletin. Most of his time was devoted to writing and teaching, although he went on a number of field trips, including visits to Iran, Afghanistan, and the American North West, the results of which fed into his seminars and into articles. Although a reclusive figure, Benveniste was in touch with many of the well-known linguists of his time. His correspondence with Hjelmslev shows a common approach to a structuralist methodology; in 1947, Benveniste became one of the few foreign members of the Copenhagen Linguistic Circle. During the intellectual ferment of the late 1960s in France, Benveniste became a slightly more public figure, giving interviews on linguistic structuralism, and being championed by such thinkers as Kristéva and Barthes. On the founding of the International Association for Semiotic Studies in Paris in January 1969, he became its first president. Benveniste had already suffered a debilitating coronary thrombosis in 1956, and he had an incapacitating stroke in 1969, having seen his book on the vocabulary of Indo-European institutions through to completion only months before. In his last difficult years of failing health, he was nursed with dedication by his sister, and his cause was promoted by contemporaries such as Barthes and Jakobsen. He died in October 1976.

References

Arrivé, Michel, and Claudine Normand (eds.) 1992. Emile Benveniste Vingt Ans Après, Linx no. 26, 1.

Benveniste, Emile. 1935. Origines de la formation des mots en indo-européen, Paris: Adrien-Maisonneuve.

———.1948. Noms d'agent et noms d'action en indo-européen, Paris: Adrien-Maisonneuve.

———.1962. Hittite et indo-européen, Paris: Adrien-Maisonneuve.

———.1971. Problèmes de linguistique générale I, 1966, as Problems in General Linguistics, transl. by M.E. Meek, Coral Gables, FL: University of Miami Press.

———. 1973. Le vocabulaire des institutions européennes, 1969, as Indo-European Language and Society, transl. by E.Palmer, London: Faber & Faber.

———. 1974. Problèmes de linguistique générale II, Paris: Gallimard.

Lotringer, Sylvère, and Thomas Gora (eds.) 1981. Polyphonic linguistics: the many voices of Emile Benveniste, Semiotica, special supplement.

Kristéva, Julia, et al. (eds.) 1975. Langue Discours Société. Pour Emile Benveniste, Paris: Seuil.

Malkiel, Yakov. 1980. Lexis and grammar — necrological essay on Émile Benveniste (1902–76). Romance Philology 34(2). 160–94.

Montaut, Annie, and Claudine Normand (eds.) 1997. Lectures d'Emile Benveniste, Linx no. spécial.

CAROL SANDERS

See also **Indo-European 1: Overview; Semiotics; Structuralism**

Bever, Thomas

Within linguistics, Thomas Bever can be considered a renaissance scholar. His work impacts the fields of theoretical linguistics, psycholinguistics, cognitive neuroscience, reading, esthetics, and language development. He is probably best known for his work that distinguishes independent systems of language processing (production and perception) from the system that represents grammatical knowledge. He argues that the universal properties of language can come from a variety of sources, including what can be said, understood, remembered, and learned, and are not necessarily the result of grammatical knowledge.

Originally interested in music and theater, Bever reports that he decided to take a major in linguistics primarily because it was the course of study with the fewest requirements that involved language courses. While an undergraduate, he worked in the psycholinguistics lab of Margaret Bulowa (assisting in an early study of language acquisition using a camera system he designed). His undergraduate thesis was on the stages of the acquisition of phoneme perception in infants. He was one of the first students in the linguistics doctoral program at MIT, where he studied both linguistics (mostly phonological theory under Morris Halle) and psychology.

In 1964, Bever was elected to the Harvard Society of Fellows, where he worked in the laboratory of George A. Miller. He continued working with Miller when they both moved to Rockefeller University in 1967. It was in Miller's lab that he first explored the question of the relationship between formal grammatical models and psychological reality. Here, he collaborated extensively with Jerry Fodor, Merrill Garrett, and Jacques Mehler. Bever's earliest work was partly a response to behaviorist claims that speakers and language learners do not have access to 'abstract' or 'deep' levels of language. Using perceptual experiments, Bever and his colleagues showed that speakers do indeed have access to abstract grammatical information. The experiments involved placing a click in a neutral place in the stream of sounds. People tend to perceive these clicks not in the place where they actually occur, but at some phrasal boundary. Bever extended these tests to even more abstract structures. The following two sentences, although apparently similar on the surface, differ in the deep semantic relations underlying them. In (2a), the noun phrase *the troops* is the patient of the main clause. In (2b), the troops is the agent of the embedded clause:

(2) (a) The general defied the t/roops$_i$ [to fight]
         #

     (b) The general desired [the t/roops to fight]
         #

(*from Bever 1998:128*)

Syntactic constituency is marked with square brackets; the placement of the click is marked with a slash. People perceive the click in different places (marked with a #) in the two sentences, corresponding to the structure of the sentence. The click experiment showed that speakers were aware of the deep semantic difference in the structure of these clauses: the click was perceived at a clausal constituent boundary reflecting the deep positioning of the noun phrase.

These click experiments led Bever to another major area of research, that of cerebral dominance (the idea that one half of the brain is more powerful) in language and other cognitive skills, both in humans and in animals.

Perhaps Bever's most important contribution to linguistics showed that there is no single system that contributes to our production and processing of language. In Generative Grammar, there are a set of rules

for generating the structure of sentences. These rules constitute our knowledge or 'competence' in language. Linguists tap this knowledge using acceptability judgments. In early work, George Miller hypothesized that the set of generative rules were essentially identical to the set of processes we use to produce and decode sentences. One consequence of this idea is that the more rules involved in generating a sentence, the harder it should be to produce or understand; this hypothesis is often known as the 'Derivational Theory of Complexity'. While initial experimental results testing this were promising, several studies by Bever and his colleagues showed that it was based on some faulty assumptions about how the rules were organized. Bever (in work that continues up to the present time) suggests that perhaps more than one system is at work when processing or producing a sentence. For example, when hearing a sentence, one uses a set of perceptual strategies to provide an initial structure to the sentence. One example of such a strategy is the idea that noun–verb–noun (NVN) sequences map onto agent–action–patient semantics; another is that the first noun–verb sequence heard constitutes the subject and predicate of the main clause. These perceptual strategies are not the same as the generative rules that give us acceptability judgments. Instead, these strategies form a separate cognitive system. Evidence for this proposal comes from the fact that at the stage when children seem to rigidly use perceptual strategies in other domains (such as in estimating numerosity)—at approximately 3 years of age—they rigidly apply linguistic perceptual strategies to sentences they hear, often doing much worse than younger children. For example, in tests where they are asked to identify which picture is best associated with a particular sentence (3), 2-and 4-year-olds both correctly identify *Big Bird* as the agent and *Cookie Monster* as the theme. However, 3-year-olds reverse the roles, consistent with the NVN strategy, which treats the first noun as an agent.

(3) Cookie Monster was hit by Big Bird

Since there is a stage at which children seem to exclusively use these perceptual strategies, it seems reasonable to conclude that they exist.

Obviously, these perceptual strategies must relate in some way to grammatical rules. Bever claims that we use both systems in a kind of 'analysis by synthesis': we use perceptual strategies to provide an initial hypothesis about the structure of the sentence and the semantic relations that underlie it, then we go back with our generative rules and check the structure. One piece of experimental evidence that he presents is that tone and click perception at the end of clauses is poor, suggesting that this is the stage at which synthesis is occurring.

Bever's wide-ranging interests reflect the complexity of the phenomena he studies. His work has raised ques-

tions about the intersections among language acquisition, sentence processing, formal grammar, and neuroscience (among other areas) and has shown that one cannot thoroughly theorize about any one of these without considering the interactions with other domains.

Biography

Thomas Bever was born in Boston, Massachusetts, on December 9, 1939. He completed his B.A. (1961) magna cum laude with highest honors in psychology and linguistics from Harvard, tutored by Roman Jakobson, and his undergraduate thesis was on the stages of the acquisition of phoneme perception in infants. His Ph.D. (1967), Massachusetts Institute of Technology, dissertation on Menomini phonology was supervised by Morris Halle. He was Lecturer, MIT Psychology department, 1964–1966; Assistant Professor, 1967–1969, and Associate Professor, 1969–1970, Rockefeller University; Professor of Psychology, Columbia University, 1970–1986; Chairman, Columbia University Interdisciplinary Ph.D. Program in Psychology and Linguistics, 1973–1986; Pulse Professor of Psychology and Professor of Linguistics, University of Rochester, 1985–1995; Head, Language and Cognition Program 1986–1989 and 1992–1994, Director, Cognitive Science Program 1991–1992, Directory, Center for the Sciences of Language 1988–1995 at the University of Rochester; Research Professor of Cognitive Science, Linguistics and Psychology, University of Arizona, 1995–present; Head, Department of Linguistics, University of Arizona 1998–2003; Phi Beta Kappa 1961, National Institutes of Health Predoctoral Fellowship 1962–1964, Harvard Society of Fellows 1964–1967; National Science Foundation Faculty Fellowship, summers 1974–1977, Guggenheim Fellowship 1976–1977, Fellow, Center for Advanced Study in the Behavioral Sciences 1984–1985; Vice-President, the Rockefeller University Chapter of the American Association of University Professors, 1969–1970; and Co-founder and Associate Editor, *Cognition*, 1973–present.

References

Berwick, Robert C., and Amy S. Weinberg. 1984. The grammatical basis of linguistic performance: language use and acquisition. Cambridge: MIT Press.

Bever, Thomas. 1970. The cognitive basis for linguistic structures. Cognition and language development, ed. by R. Hayes, 279–362. New York: Wiley & Sons, Inc.

Bever, Thomas. 1980. Broca and Lashley were right: cerebral dominance is an accident of growth. Biological studies of mental processes, ed. by D. Caplan, 186–230. Cambridge, MA: MIT Press.

Bever, Thomas. 1988. The psychological reality of grammar: a student's eye view of cognitive science. The making

of cognitive science: a festschrift for George A. Miller, ed. by W. Hirst, 112–42. Cambridge: Cambridge University Press.

Bever, Thomas, J.M. Carroll, and L.A. Miller (eds.) 1984. Talking minds. Cambridge, MA: MIT Press.

Bever, Thomas, and J. Fodor. 1965. The psycholinguistic reality of linguistic segments. Journal of Verbal Learning and Verbal Behavior 4. 414–20.

Bever, Thomas, and B. McElree. 1988. Empty categories access their antecedents during comprehension. Linguistic Inquiry 19.

Bever, Thomas, H.L. Roitblat, and H.S. Terrace (eds.) 1984. Animal cognition. Hillsdale, NJ: Erlbaum.

Bever, Thomas, M. Sanz, and D. Townsend. 1998. The emperor's psycholinguistics. Journal of Psycholinguistic Research 27. 261–84.

Bever, Thomas, and D. Townsend. 2001. Sentence processing. Cambridge, MA: MIT Press.

Carroll, J.M., Bever, Thomas, and C.R. Pollock. 1981. The non-uniqueness of linguistic intuitions, Language 57. 368–82.

Carroll, David W. 1986. Psychology of language. Pacific Grove, CA: Brooks/Cole Publishing Company.

Fodor, Jerry, Thomas G. Bever, and Merrill Garrett. 1974. The psychology of language: an introduction to psycholinguistics and generative grammar. New York: McGraw-Hill.

Garrett, Merrill. 1990. Sentence processing. An invitation to cognitive science. Vol 1: Language, ed. by Daniel N. Osherson and Howard Lasnik, 133–75. Cambridge, MA: MIT Press.

Schütze, Carson. 1996. The empirical base of linguistics grammaticality judgments and linguistic methodology. Chicago: The University of Chicago Press.

ANDREW CARNIE

Bilingual Acquisition

Children growing up exposed to two or more languages acquire each of them in much the same way as monolinguals (speakers of just one language). Bilingual acquisition during childhood can thus be regarded as an instance of simultaneous acquisition of two 'first' languages. In fact, if children are exposed to more than two languages simultaneously, they are able to acquire full competence of each that does not substantially differ from the speaking ability of monolinguals; bilingualism is therefore a special case of multilingualism.

Comparison with monolinguals is one of the main issues in research on bilingual acquisition. It can be extended to all types of acquisition where more than one language is learned, independently of age of onset and proficiency attained in each language. The present discussion, however, is limited to child bilingualism. Whether types of bilingual acquisition should, in fact, be distinguished according to the age of onset of learning ('bilingualism' vs. 'foreign language learning') depends ultimately on the results obtained in these comparisons. Given our current state of knowledge, it is plausible to assume that age of onset is a crucial factor causing fundamental differences between child and adult language acquisition. Consequently, it is necessary to distinguish between simultaneous and successive acquisition of bilingualism. 'Simultaneous' acquisition does not necessarily imply that the child has been exposed to both languages from birth, although this interpretation has

occasionally been suggested. The controversial issue here is the age range during which changes occur, which result in qualitative differences between first and second language acquisition. Although sufficient evidence is not yet available, the period around three years of age appears to be of particular significance. The most widely accepted view is that one may qualify as 'simultaneous' those instances of bilingual acquisition where the child is exposed to two languages before the age of three. In this case, each language of the bilingual is predicted to share crucial similarities with that of the respective monolingual speakers but to exhibit differences in comparison with languages learned at a later age.

The insight that bilingual acquisition proceeds in the same way and leads to the same type of linguistic knowledge as monolingual acquisition is largely due to a number of case studies carried out in the late 1970s and 1980s, although the first monograph documenting the linguistic development of a bilingual child (Ronjat 1913) arrived at a similar conclusion. Until the 1970s, it was frequently assumed that bilinguals face serious difficulties in separating language systems and that they run the risk of ending up with mixed languages. Such views were often the result of inadequate research methodologies. Large-scale studies, frequently focusing on disfavored minority groups, failed to perform in-depth analyses of bilingual language use. Early case studies, carried out by amateurs without the appropriate training, analyzed the language use of

their own children. Some more satisfactory scholarly studies, however, arrived at similar conclusions, mainly because they adopted a monolingual perspective in their investigation of bilingualism. The resulting shortcomings were amended by later research when, beginning in the 1980s, interest in bilingual acquisition increased dramatically.

Other issues that have attracted the attention of both researchers and the interested public involve alleged problems in separating the two languages. Bilinguals tend to mix languages within a conversation or even an utterance; from a monolingual perspective, this appears to indicate an inability to keep the languages apart. It has also been suggested that bilingual children initially go through a phase during which they develop one 'hybrid' system before they are able to differentiate the lexical and grammatical systems they are acquiring. As early as the 1970s, researchers agreed almost unanimously that these children eventually succeed in separating their languages, without much effort or specific pedagogical measures, but parents and educators tended to regard the alleged single linguistic system phase as a potentially serious problem, fearing the possibility of a lasting state of incomplete competence in both languages. More sophisticated analyses of children's language mixing eventually revealed no initial single-system stage. Children have been shown to differentiate linguistic systems from very early on. Grammatical differentiation, for example, is evidenced as soon as multiword utterances are used. Thus, the available empirical evidence does not support the idea of a temporary fusion of grammatical systems, and mixing can for the most part be explained as a particular type of bilingual language use. Early grammatical differentiation, however, does not exclude the possibility of crosslinguistic influence during later developmental phases, resulting either in vocabulary or grammatical transfer or in acceleration or delay of development. The question of whether such influence actually plays a major role in bilingual acquisition continues to be controversial.

Language mixing by adult bilinguals is normally the result of code-switching, a form of language use determined by a complex network of sociolinguistic factors. Code-switching within a single clause is subject to grammatical constraints depending on the structural properties of both languages. Consequently, in order to be able to code-switch, children have to acquire the required social skills and grammatical knowledge. It has therefore been suggested that mixing during initial phases of linguistic development may not yet be adult-like code-switching. However, before the age of 2, children choose the language according to the addressee and begin to comply with

other sociolinguistic requirements. No later than during the first half of the third year, mixes within a single clause conform to grammatical constraints. Since earlier examples consist almost exclusively of single-word mixes, the role of structural constraints is difficult to assess. The frequency of mixing does not generally decrease with grammatical development, contrary to attempts to explain early mixing as resulting from deficient ability. Although not knowing a word in one language may initially lead to mixing, it is also triggered by social context. Most importantly, the language children are exposed to appears to influence their mixing behavior. Not only do children mix more frequently if the persons they interact with mix freely, but also the nature of parent's reaction to child mixing has been argued to be of crucial importance, i.e. parents signal overtly or implicitly whether they encourage or discourage language mixing.

Quantitative and qualitative characteristics of language exposure are also relevant for the question if bilinguals can attain balanced, i.e. equally developed, knowledge of both languages. One language might tend to be dominant (preferred choice, faster recall of words, etc.), but dominance can shift over the lifespan, and it does not seem to affect competence—no convincing evidence supports the idea of 'semicompetence'. Similarly, although it has been suggested that bilingual acquisition is delayed in comparison to monolingual acquisition, onset as well as development are clearly within the limits of monolingual norms, faster bilinguals outperforming slower monolinguals. In both cases, differences concern details of speaking, not qualitative differences in underlying knowledge. This confirms the hypothesis that the human language faculty predisposes the individual to become bilingual.

References

De Houwer, Annick. 1990. The acquisition of two languages from birth: a case study. Cambridge: Cambridge University Press.

Genesee, Fred. 1989. Early bilingual development: one language or two? Journal of Child Language 16. 161–79.

Grosjean, François. 1982. Life with two languages: an introduction to bilingualism. Cambridge, MA and London: Harvard University Press.

Köppe, Regina, and Jürgen M. Meisel. 1995. Code-switching in bilingual first language acquisition. One speaker, two languages: cross-disciplinary perspectives on code-switching, ed. by Lesley Milroy and Pieter Muysken, 276–301. Cambridge: Cambridge University Press.

Lanza, Elizabeth. 1997. Language mixing in infant bilingualism: a sociolinguistic perspective. Oxford: Clarendon Press.

Leopold, Werner F. 1939–1949. Speech development of a bilingual child: a linguist's record, vols. 1–4. Evanston, IL: Northwestern University Press; New York: AMS Press, 1970.

Meisel, Jürgen M. 1989. Early differentiation of languages in bilingual children. Bilingualism across the lifespan: aspects of acquisition, maturity, and loss, ed. by Kenneth Hyltenstam and Loraine K. Obler, 13–40. Cambridge: Cambridge University Press.

Paradis, Johanne, and Fred Genesee. 1996. Syntactic acquisition in bilingual children: autonomous or interdependent? Studies in Second Language Acquisition 18. 1–25.

Ronjat, Jules. 1913. Le développement du langage observé chez un enfant bilingue. Paris: Librairie Ancienne H. Champion.

Volterra, Virginia, and Traute Taeschner. 1978. The acquisition and development of language by bilingual children. Journal of Child Language 5. 311–26.

JÜRGEN M. MEISEL

See also **Acquisition; Acquisition Theories**

Bilingual Mixed Languages

It is generally believed that a language's basic vocabulary and grammatical inflections are relatively impervious to foreign influence, and that they are therefore safe diagnostic tools for determining whether given languages are related. It is also said that 'no language is so mixed that it cannot fit unambiguously into a family tree: it will always be possible to show that the bulk of a language's lexicon and grammatical structures come from the same source.' According to these claims, therefore, structurally mixed languages should not exist.

Nevertheless, mixed languages—languages that cannot be classified unequivocally—do in fact exist and are relatively common. The best-known examples are languages that arise as a direct result of language contact and that comprise linguistic material that cannot be traced back to a single source language. Such languages therefore do not belong to any language family and their genesis is not a matter of descent with modification from a single parent language. It is generally accepted that mixed languages are of three types: pidgins, creoles, and bilingual mixtures (also known as 'intertwined languages'). In pidgins and creoles, the vocabulary is usually traceable to a single language, always the language of the dominant group involved in the contact, while the grammatical structure (and often the sound system) generally cannot be traced to any single language. In bilingual mixed languages, the basic vocabulary can always be traced to a single source language and the grammatical structure (including the sound system) to another single source language. Furthermore, the linguistic material from each source language is adopted wholesale without the kind of distortion that occurs in pidgins or creoles. Also bilingual mixtures share a characteristic social feature: all of them are in-group languages used within the community as a sign of community solidarity and are not understood by outsiders, which explains why they have often remained undetected. Any type of language can be involved in bilingual mixed languages, and there are no structural constraints that prevent the 'intertwining' of any two languages, although the result may differ according to the typological properties of the languages involved.

Media Lengua, spoken in Ecuador, is an example of a bilingual mixed language. The vocabulary is of Spanish origin and the grammar stems from Quechua, a Native American language, as the following diagram for the sentence 'I come to ask a favor' shows:

Media Lengua (Ecuador): Spanish lexicon—Quechua grammar; Spanish items are in italics. Source: Muysken (1997).

ML	*Unu fabur* - ta	*pidi* - nga - bu
	ONE FAVOR - ACC	ASK - NOM - BEN
	bini - xu - ni	
	COME- PROG- 1P	
Q	Shuk fabur - da	mãa - nga - bu
	ONE FAVOR-ACC	ASK- NOM - BEN
	shamu - xu - ni	
	COME - PROG - 1P	
S	Vengo para pedir un favor	
	COME 1SG FOR ASK-INF A FAVOR	
	'I come to ask a favor'	

Thus, Media Lengua uses the Quechua grammar as a structural grid into which the Spanish vocabulary items are inserted.

Michif, spoken in Western Canada and in North Dakota, represents a similar case. Here Cree, a Native American language, provides the structural basis for the insertion of French vocabulary.

Other examples of bilingual mixed languages are:

Source of Vocabulary		Mixed Language		Source of Grammar
Puquina	→	Callahuya (Bolivia)	←	Quechua
Cushitic	→	Ma'a / Inner Mbugu (Tanzania)	←	Bantu
Japanese	→	Senkyoshigo (Japan)	←	English
Dutch	→	Krõjo / Javindo (Java)	←	Low Javanese
Dutch	→	Pecu' (Surabaja, Indonesia)	←	Low Javanese
Malay	→	Chindo (Java)	←	Low Javanese
Cushitic	→	Ilwana (Kenya)	←	Bantu

Angloromani (Great Britain, USA, Australia): Romani lexicon—English grammar; Romani items are in italics. Source: Bakker and Muysken (1994).

> AR *Palla bish besh* - es *apoli* the *Beng*
> AFTER 20 YEAR – PLUR AGAIN DEVIL
>
> *wel* - d and *pen*-d : *Av* with *man- di*
> COME-PAST SAY-PAST COME ME – DAT

> R Palla bish besh- aw apopli o Beng vi-as. Yov pen-das : Av man-tsa

> E After twenty years the Devil came back and said : Come with me

Over a dozen Romani mixed languages are spoken by Gypsies in Europe and the Middle East. In all cases, the vocabulary is Romani and the grammar comes from the language of the host country (Basque Romani, Spanish Romani, Swedish Romani, Armenian Romani, Turkish Romani, etc.).

The process responsible for such bilingual mixtures has been called 'relexification', where the entire vocabulary or 'lexicon' is borrowed into another language. Conversely, 'regrammaticalization', the intrusion of an entire grammar onto a native lexicon, is equally possible. This seems to be true for the Romani mixed languages, since it is most likely that the language of the host country was learned while maintaining the original Romani vocabulary. This is why it is important to distinguish the linguistic process from the historical or social process of bilingual language mixtures. Linguistically speaking, it is generally the intrusion of a lexicon into a grammatical system (relexification), whereas the sociohistorical process may go either way.

Two specific circumstances for the creation of bilingual mixtures have been proposed. First, the birth of a new ethnic group comes about after a first generation with mothers who speak one language and fathers who speak another. The mixed language will have the grammatical system of the mothers' language and the vocabulary of the fathers' language. Second, a group of nomadic people settle and eventually feel the need for a secret trade language. They will tend to use the grammatical system of the host society and the lexicon of the inherited language. In each case, the language known best provides the grammar. The group creating the mixed language must be highly bilingual when the language mixing starts. Also, the resulting language is always intended as an in-group language and therefore is not intended to bridge a communication gap between speakers of different languages, in contrast to pidgins and creoles.

References

Arends, Jacques, Pieter Muysken, and Norval Smith (eds.) 1994. Pidgins and Creoles: an introduction. Amsterdam and Philadelphia, PA: Benjamins.

Bakker, Peter. 1996. Language intertwining and convergence: Typological aspects of the genesis of mixed languages. Sprachkontakt und Grammatikalisierung, ed. by Nicole Nau and Martin Haase, Special Issue of Sprachtypologie und Universalienforschung 49(1).

Bakker, Peter. 1997. A language of our own : the genesis of Michif, the mixed Cree–French language of the Canadian Métis. New York and Oxford: Oxford University Press.

Bakker, Peter. 2000. Rapid language change: creolization, intertwining, convergence. Time depth in historical linguistics, ed. by Colin Renfrew, April McMahon, and Larry Trask. Cambridge: McDonald Institute for Archeological Research.

Bakker, Peter, and Hein van der Voort. 1991. Para-Romani languages: an overview and some speculations on their genesis. In the margin of romani: gypsy languages in contact, ed. by Peter Bakker and Marcel Cortiade. Amsterdam: Publications of the Institute for General Linguistics, University of Amsterdam.

Bakker, Peter, and Maarten Mous (eds.) 1994. Mixed languages: Fifteen case studies in language intertwining. Amsterdam: Institute for Functional Research into Language and Language Use, University of Amsterdam.

Bakker, Peter, and Pieter Muysken. 1994. Mixed languages and language intertwining In Arends, Muysken, and Smith.

Bakker, Peter, and Robert A. Papen. 1997. Michif: a mixed language based on Cree and French. In: Thomason.

Muysken, Pieter. 1981. Halfway between Quechua and Spanis: the case for relexification. Historicity and variation in Creole studies, ed. by Arnold Highfield and Albert Valdman. Ann Arbor, MI: Karoma.

Muysken, Pieter. 1997. Media Lengua. In Thomason.

Smith, Norval. 1994. An annotated list of Creoles, Pidgins and mixed languages. In Arends, Muysken, and Smith.

Thomason, Sarah G. (ed.) 1997. Contact languages: a wider perspective. Amsterdam and Philadelphia, PA: Benjamins.

Thomason, Sarah G., and Terrence Kaufman. 1988. Language contact, creolization and genetic linguistics. Berkeley, CA: University of California Press.

ROBERT A. PAPEN

See also **Chinese Pidgin English; Chinese Pidgin Russian; Pidgins and Creoles**

139

Bilingualism

Broadly defined, bilingualism denotes the ability to use two languages. This can involve the ability to read, write, comprehend, or speak to varying levels of fluency in each language. There is a wide degree of variation across bilingual speakers in terms of their language skills. This variation stems from the manner in which the languages were acquired, the means by which they were learned, and the frequency of their use. The age of acquisition and degree of exposure to languages will also determine the degree of bilingualism.

In most cases, bilingual speakers have one dominant language, that is, they are more proficient in one or more linguistic levels of language processing in one language than in the other. In some cases, the dominant language is also the native or first language spoken. The second language is often termed 'nondominant' or 'subordinate'. Individuals can be equally proficient in two languages and are then called 'balanced bilinguals'. When individuals learn both languages simultaneously, typically from birth, they are said to be 'compound bilinguals'. In other cases, individuals learn one language first, and a second language in a different context, such as in a different country or with a different parent. The latter bilinguals are often referred to as 'coordinate bilinguals'. When individuals learn a language in this way and the first language remains relatively intact, the term 'additive bilingualism' is used. Interestingly, an individual's second language can sometimes develop into the more dominant language and actually replace the first language learned. For example, an individual born in France who is fluent in French may travel to the United States at a very young age and remain there for many years. This individual may come to know and use English with greater fluency than their native language, French. This type of bilingualism has often been referred to as 'subtractive' and may actually involve the attrition of skills in the earlier language.

Language ability can be tied to specific topics or domains of knowledge that a bilingual possesses. For example, an individual might use Spanish at home and in conversation with friends, but English at work or for technical discussion of a specific topic. Some individuals may be native speakers of a language such as Chinese and have a strong or native-like proficiency in another language, such as English, yet perform all mathematical functions in their native language only. These speakers claim that they have more automatic processing of mathematical concepts in the language

in which they originally learned them. In this way, language use may be context-specific despite a high level of fluency in more than one language. In summary, different sociolinguistic situations may determine the language that is primarily used and the manner in which it is used.

Code-switching, or the alternating use of more than one language, is common among bilingual speakers. Individuals may either mix languages within a statement or switch at defined points such as breaks between topics. While the specific reasons for switching are not yet clearly understood, it is known that bilingual speakers often switch languages in order to facilitate communication or to express concepts that can be language-specific. In this way, code-switching serves a pragmatic or strategic function. Code-switching is a complex, rule-governed phenomenon that is guided by the particular topic, context, or circumstance regarding the communication at hand. The term 'Spanglish' has been accorded to the intermixing of the Spanish and English languages in certain areas within the United States where both languages are often spoken (e.g. Southwestern states, New York, New Jersey, Florida). Researchers view code-switching as advantageous in some situations as it helps to preserve and maintain both languages simultaneously. Others argue that the constant mixing of languages may make it difficult for individuals to know any single language with a high degree of proficiency. Nevertheless, the use of languages in an interrelated way is common among individuals who share a diverse but common language background.

It is agreed that bilingualism is much more the norm around the world than monolingualism. In Indonesia, for example, several hundred languages are in constant use. In Central Europe, the ability to speak more than one language is often typical. Moreover, it has been documented that bilingual speakers possess a certain cognitive or mental flexibility that serves as an asset in cognitive processing in general. For example, bilingual children outperform their monolingual counterparts in tasks that involve role-playing, classifying objects, creativity, concept formation, memory, metalinguistic awareness, perceptual disembedding, problem-solving, role-taking, social sensitivity, and understanding complex instructions. High-performing five- and six-year-old bilingual speakers exhibit a greater degree of divergent

thinking, imagination, grammatical awareness, and perceptual organization. Clearly, bilingual children have more than one way of labeling their environment and have an early awareness of the multiple ways in which concepts can be described and named across languages. Bilingual children simply have more examples for words and grammar to choose from than do monolinguals. Therefore, language competence often influences the range of interactive skills and capabilities that a bilingual can use.

A distinction can also be drawn between 'individual bilingualism' (the bilingual competence of a single individual) and 'societal bilingualism' (the degree to which two languages are used within a society). In some cultures and language groups, bilingualism is widely accepted and languages are almost interchangeable across contexts. In other societies, multiple languages may be used, but they may be directly linked to specific contexts or uses. Diglossia is the term used to describe two functionally different varieties of the same language that are used for specific purposes. For example, some documents may be written in a more formal or Classical Arabic that is distinct from a colloquial Arabic that is more commonly used in conversation.

Other terms that are often used to refer to bilingualism include 'polyglottism' (although 'polyglot' often refers to people who allegedly command an improbably large number of languages) and 'multilingualism'.

References

Altarriba, J. 2000. Language processing and memory retrieval in Spanish–English Bilinguals. Spanish Applied Linguistics 4. 215–45.

Baker, C. 1988. Key issues in bilingualism and bilingual education. Clevedon: Multilingual Matters.

Bialystok, E. (ed.) 1991. Language processing in bilingual children. Cambridge: Cambridge University Press.

de Groot, A. M. B., and J. F. Kroll (eds.) 1997. Tutorials in bilingualism. Mahwah, N J: Lawrence Erlbaum Associates.

Grosjean, F. 1982. Life with two languages. Cambridge, MA: Cambridge University Press.

Hakuta, K. 1986. Mirror of languages: the debate on bilingualism. New York: Basic Books.

Harris, R. J. (ed.) 1992. Cognitive processing in bilinguals. Amsterdam: Elsevier.

Schreduer, R., and B. Weltens (eds.) 1993. The bilingual lexicon. Philadelphia: John Benjamins.

Weinreich, U. 1953. Languages in contact. New York: Linguistic Circle of New York.

JEANETTE ALTARRIBA

See also **Bilingual Acquisition**

Biosemiotics

Biosemiotics deals with all kinds of sign-based processes in living systems. In other words, it is a theory of semiosis in living systems, or biology that interprets living systems as sign systems; or the study of biological codes; or the study of prelinguistic sign processes and signs. These definitions can be taken as (almost) synonymous. The appearance of human language does not mean the primary appearance of signs, because there are signs preceding human language. Accordingly, human language comprises a special type among the sign systems. There exists the *semiotic threshold* between living and nonliving, a natural border between the semiotic and nonsemiotic world, which means that there are no sign processes *per se* in the nonliving world, whereas life is perfused with signs, if it is not composed exclusively of signs. According to a biosemiotic view, life science and sign science are coextensive.

Altogether, there exist five distinctively different levels of sign systems.

(1) *Cellular sign systems*. This is the sign system of any bacterial (and eukaryotic) cell. Its characteristic processes are enzymatic recognition and transmembrane signaling. Its inheritance system combines the epigenetic and genetic, the latter being dominant. It is characterized by microsemiosis. According to Yates (1997:458), microsemiosis 'does not address communication between cells or among cell complexes'. In a cell, true *codes* appear. The codes, as different from other relationships, are the relationships that are not deducible from universal physical laws because they connect independent domains and are fixed due to a unique historical process (Barbieri 2001). Codes as part of any single semiosis turn the web-like set of cellular processes into a meaning-generating interpretation-like process—this is biointerpretation, or biosemiosis.

(2) *Vegetative sign systems*. This includes the communication between the tissues in a multicellular organism. Because the cells of an organism are genetically identical, the differences between the cell types of an organism are based on another inheritance system—the epigenetic one. Vegetative sign systems are responsible for the genesis of a multicellular biological form, the whole morphology of the body as a result of communication between cells (Kull 2000). Morphogenesis and cell differentiation are its typical representatives. In this sense, vegetative sign systems are not confined to plants—they occur in all multicellular organisms. Due to them, organic forms are communicative structures. Vegetative sign systems provide categorization without representation.

(3) *Animal sign systems*. This is the senso-neuro-muscular system, the one that is responsible for the behavior of a motile animal organism. The basic inheritance system here is neuro-humoral (or behavioral). Its characteristic feature is the existence of representations, and thus—Umwelt, the model of the organism's world. Animal sign systems represent the primary modeling systems (according to Sebeok 1994) and are studied by zoosemiotics.

(4) *Linguistic sign systems*. This corresponds to the secondary modeling system (natural language), according to Sebeok (1994) (or primary modeling systems, by Lotman). Different from the animal sign system, it has syntactic signs (Bickerton 1990), i.e. the signs that are used to refer to a type of relation between signs (e.g., a human 'and'—a sign which other animals lack), and is principally symbolic (Deacon 1997). Its characteristic feature is the existence of sentences and thus symbolic language.

(5) *Cultural sign systems*. This corresponds to the tertiary modeling systems, according to Sebeok (or secondary modeling systems, according to Lotman). Their characteristic features include the existence of artistic, ideological, ethical, and similar structures.

When characterizing the semiotic aspects of behavior and communication of different groups of animals (e.g., insects, birds, mammals), it is important to pay attention to (1) the general features of the body plan and sensory apparatus of the animals, and the functional place of the animal group in ecosystems, (2) the main ways and means of active communication, and (3) the main trends in the evolution of communication in the group.

Because nonhuman communication does not use sentences, that is, it is nonpropositional, the entire emphasis of the study in biosemiotics appears to be very different from the one in linguistic sciences. However, biosemiotic knowledge is important for linguistics due to its results in the analysis of more fundamental and general communication phenomena. For instance, the general mechanism of categorization, as known in phonemics, has been shown to be analogous to the mechanisms of speciation in biparental organisms.

A pioneering approach to biosemiotics has been established by the studies of a biologist and philosopher Jakov von Uexküll (1928, 1982, 1992). In the framework of semiotics, biosemiotics has been developed basically together with a turn from a Saussurean dyadic to Peircean triadic concept of sign (Deely 2001) and with a development of concept of semiosphere as a web of all kind of (including interspecific) communication processes. A key concept in it is an analog-digital duality of all organic meaning-generating processes (Hoffmeyer, Emmeche 1991). Since 1990s, biosemiotics has established as a university discipline (Sebeok, Umiker-Sebeok 1992, Kull 2000, Sebeok 2001, Schult 2004).

References

Barbieri, Marcello. 2001.The organic codes: the birth of semantic biology. Ancona: Pequod.

Bickerton, Derek. 1990. Language and species. Chicago: University of Chicago Press.

Deacon, Terrence. 1997.The symbolic species: the co-evolution of language and the brain. NY: Norton.

Deely, John. 2001. Umwelt. Semiotica 134(1/4). 125–35.

Hoffmeyer, Jesper, and Emmeche Claus. 1991. Code-duality and the semiotics of nature. On semiotic modeling, ed. by M. Anderson and F. Merrell, 117–66. Berlin: Mouton de Gruyter.

Kull, Kalevi. 1999. Biosemiotics in the twentieth century: a view from biology. Semiotica 127(1/4). 385–414.

Kull, Kalevi. 2000. An introduction to phytosemiotics: semiotic botany and vegetative sign systems. Sign Systems Studies 28. 326–50.

Schult, Joachim (ed.) 2004. Biosemiotik: praktische Anwendung und Konsequenzen für die Einzelwissenschaften. Berlin: Verlag für Wissenchaft und Bildung.

Sebeok, Thomas A. 1994. Signs: an introduction to semiotics. Toronto: University of Toronto Press.

Sebeok, Thomas A. 2001. Biosemiotics: its roots, proliferation, and prospects. Semiotica 134(1/4). 61–78.

Sebeok, Thomas A., and Jean Umiker-Sebeok (eds.) 1992. Biosemiotics. Berlin: Mouton de Gruyter.

Uexküll, Jakob von. 1928. Theoretische Biologie. Berlin: Springer.

Uexküll, Jakob von. 1982. The theory of meaning. Semiotica 42(1). 25–82.

Uexküll, Jakob von. 1992. A stroll through the worlds of animals and men: a picture book of invisible worlds. Semiotica 89(4). 319–91.

Yates, F. Eugene. 1997. Microsemiosis. Semiotics: a handbook on the sign-theoretic foundations of nature and culture, Vol. 1, ed. by Roland Posner, Klaus Robering and Thomas A. Sebeok, 457–64. Berlin: Walter de Gruyter.

KALEVI KULL AND ALEKSEI TUROVSKI

Bloomfield, Leonard

Leonard Bloomfield may be considered the father of modern American linguistics. His masterwork on general linguistics, *Language* (1933), determined the direction the scientific study of language in the United States would take for the next three decades. A prolific scholar, Leonard Bloomfield's publications cover an impressive spectrum: general linguistics, Indo-European and Germanic philology, dialect geography, linguistic descriptions of the Algonquian American-Indian languages Menomini, Cree, Fox, and Ojibwa, and the Austronesian languages Tagalog and Illocano. Bloomfield's work spawned a generation of American scholars, the neo-Bloomfieldians, who dominated the field of linguistics before Noam Chomsky's theory of transformational grammar became widely accepted in the early 1960s: Bernard Bloch, Robert A. Hall Jr., Zellig Harris, Einar Haugen, Charles F. Hockett, Martin Joos, William Moulton, Eugene Nida, and W. Freeman Twadell, to name but a few.

An adherent of positivism and behaviorist psychology, Bloomfield rejected mentalism, which takes into account nonphysical and hence nonobservable thought processes, insisting that only observable manifestations of language could be the subject of linguistic investigation. He defined the meaning of a linguistic form as the situation in which the speaker utters it and the response that it calls forth in the hearer. The strong rejection of the mentalist approach to the study of meaning became a defining characteristic of Bloomfieldian linguistics and its weakest point, as Noam Chomsky would point out several decades later. A strong believer in the scientific method, Bloomfield argued that language should be studied like a natural science. In *A Set of Postulates for the Science of Language* (1925), he outlined the assumptions underlying structural linguistics in a concise and compressed form.

As a historian of language, Bloomfield studied his subject from the diachronic perspective; as a descriptive structural linguist, he concentrated on the synchronic perspective. A diachronic approach traces the changes a language undergoes from a given point in time to a later date; a synchronic approach describes a language at one particular time. This distinction had not been made by earlier generations of scholars, and Bloomfield, like the French structuralist Ferdinand the Saussure, insisted that both perspectives be kept strictly separate. The classic diachronic approach to linguistic analysis, the comparative method, was pioneered by the *neogrammarians*, a group of nineteenth-century German scholars who established the genetic relationship among Indo-European languages. Cognates (shared vocabulary) are compared to establish the sound changes marking the transition from a common hypothetical ancestor to one or more daughter languages. A comparison of the median consonants in Old English */brothor/,* Old Saxon */brothar/,* and Old High German */bruoder/,* for example, would have led scholars to the conclusion that */th/* was present in proto-Germanic, the shared ancestral language, but changed to */d/* in Old High German. The comparative method rests on the assumption that sound change affects a language without exception and is therefore regular, never occurring sporadically in just a few words. Using the neo-grammarian principle as a working hypothesis, Bloomfield applied the comparative method to Cree, Fox, Menomini, and Ojibwa, languages some European scholars of the time might have considered primitive compared to Sanskrit, Gothic, Latin, or Greek, and proved that they were descended from a common ancestor, proto-Algonquian. In spite of criticism, especially from abroad, Bloomfield remained a confirmed neo-grammarian throughout his career.

Bloomfield's masterwork *Language* introduced innovative methods for descriptive synchronic linguistics and a new terminology soon to be adopted by his American colleagues. The first half of the book is a state-of-the-art presentation of the methods and principles of American structuralism; the second presents the findings of European historical linguists and dialectologists supplemented with Bloomfield's own research materials. He introduces the American reader to the phonemic principle established by the Prague School of Linguistics. Phonemics is the study of the distinctive sounds of a language, those that are recognized by the native speaker as signaling differences in meaning. The initial consonants of *pin, kin, sin, tin* are phonemes of English since their substitution results in a change of meaning. The different articulations of the phoneme */p/* in *pit* (aspirated), *spit* (unaspirated), and *tip* (unreleased), on the other hand, are phonetic variants of the same phoneme, which can never serve to distinguish meaning since their phonetic shape depends on their position in the word. The phonemic principle was the basis for all work in descriptive synchronic linguistics by Bloomfield and his disciples. In *Language,* the synchronic phonemic principle is, for the first time, equated with the diachronic neo-grammarian principle of the

regularity of sound change. American reviewers of *Language* were generally enthusiastic, whereas European scholars tended to be more critical citing the lack of abstract theory, rejection of mentalism, new terminology (sememe, taxeme, glosseme, etc.), and Bloomfield's defense of the neo-grammarian principle. One reviewer considered the work to be typically American: data-oriented, dryly factual, but carefully and systematically executed. In the United States, *Language* became the preferred textbook in introductory linguistics courses for the next quarter of a century following publication. It should be recommended reading for graduate students in linguistics interested in the history of their discipline.

Bloomfield began working with non-European languages early in his career. *Tagalog Texts with Grammatic Analysis* (1917) filled a void in a much neglected field. Based on the data provided by a single informant who dictated texts, it is still considered to be one of the best sources of information on any Austronesian language. While the Algonquian languages had been studied systematically for three centuries, Bloomfield's contribution was the systematic synchronic description of four languages, Cree, Fox, Menomini, and Ojibwa, and the application of the diachronic comparative method to the reconstruction of the hypothetical ancestor, proto-Algonquian. Bloomfield provided a grammatical outline of each of the four living languages as well as of proto-Algonquian, and traced the historic development of each language, thus demonstrating the validity of the comparative method in the process. He also produced the first systematic description of the Algonquian morphology (the study of inflection and word formation), and thereby established the framework for all later studies on the subject. Bloomfield's field methods and methods of linguistic analysis have stood the test of time, and are still used by modern scholars recording unwritten languages or dialects. The bulk of his Algonquian material was published posthumously by Charles F. Hockett, his literary executor.

During World WarII, Bloomfield, like many of his colleagues and students, devoted much of his time to the development of a new foreign language teaching methodology for the War Department. The audiolingual method focused on the spoken, colloquial form of the language rather than the literary standard and introduced practice sessions where the material was transmitted to the learner on the basis of inductive, behaviorist learning principles (oral repetition and memorization). By the 1960s, it had become the almost universally accepted foreign language teaching methodology in the United States and to some extent in Europe. Bloomfield authored and coauthored text books for elementary Dutch and Russian (1944/1945).

Bloomfieldian structuralism with its focus on phonemics and morphology remained the dominant approach to linguistic analysis until the publication of Noam Chomsky's treatises on transformational generative grammar in the late 1950s and early 1960s. After several decades of relative obscurity, Bloomfield's contribution to the field of linguistics is once again appreciated. Leonard Bloomfield suffered a debilitating stroke in 1946, which abruptly terminated his brilliant career and led to his death three years later.

Biography

Leonard Bloomfield was born on April 1, 1887 in Chicago, IL. The son of Austrian immigrants and the nephew of Maurice Bloomfield, a noted scholar of Indo-European philology and Indic studies, Leonard Bloomfield grew up in Chicago and rural Wisconsin.

Fluent in both English and German, he quite naturally gravitated toward the study of Germanic and Indo-European philology. After completing his undergraduate degree at Harvard in 1906, he began graduate studies in German at the University of Wisconsin (1906–1908) and then transferred to the University of Chicago where he received his Ph.D. degree in 1909 with a thesis in historical Germanic linguistics ('A semasiological differentiation in Germanic secondary Ablaut'). After several years of working as an instructor of German at the Universities of Cincinnati (1909) and Illinois Champaign-Urbana (1910–1913), he spent a year, between 1913 and 1914, of advanced study in Indo-European and Germanic philology at the German Universities of Leipzig and Göttingen, where he worked with Karl Brugner, August Leskien, and Jakob Wackernagel, some of the most highly regarded historical linguists of the time. By 1914, he was Assistant Professor of comparative philology and German at the University of Illinois. For the rest of his career, Bloomfield remained an adherent of their particular view of language change. One of the founding members of the *Linguistic Society of America* (1924) with George M. Bolling and Edgar H. Sturtevant, Bloomfield held the chair of Germanic Philology at the University of Chicago from 1927 until 1940 when he accepted a Sterling professorship in Germanic Languages at Yale University. He became president of the *Linguistic Society of America* in 1936. Bloomfield suffered a debilitating stroke in May 1946, which ended his career. He died on April 18, 1949.

References

Bloch, Bernard. 1949. Leonard Bloomfield. Language 25. 87–98.

Bloomfield, Leonard. 1917. Tagalog texts with grammatical analysis, 3 vols. Urbana: University of Illinois.

———.1925. Why a linguistic society? Language 1. 1–5.

———.1926. A set of postulates for the science of language. Language 2. 153–64.

———.1927. The word stems of Central Algonquian. Festschrift Meinhoff. Hamburg: L. Friedrichsen. 303–403.

———.1928. Menomini texts. Publication of the American Ethnological Society, Vol. 12. New York: G.E. Stechert.

———.1930. Linguistics as a science. Studies in Philology 27. 553–7.

———.1933. Language. N Y: Holt, Rinehart and Winston.

———.1942. Outline guide for the practical study of foreign languages. Baltimore: Linguistic Society of America, Special Publication.

———.1942. Outline of Ilocano syntax. Language 18. 193–200.

———.1945. About foreign language teaching. Yale Review 34. 625–41.

———.1962. The Menomini language. New Haven: Yale University Press.

———.1970. A Leonard Bloomfield anthology, ed. by Charles F. Hockett. Bloomington: Indiana University Press.

Goddard, Ives. 1987. Leonard Bloomfield's descriptive and comparative studies of Algonquian. In Hall. 179–217.

Hall, Robert A. Jr. 1987. Leonard Bloomfield: essays on his life and work. Philadelphia: John Benjamins.

Hall, Robert A. Jr. 1990. A life for language. A bibliographical memoir of Leonard Bloomfield. Amsterdam Studies in the Theory and History of Linguistic Science, Vol. 55. Amsterdam/Philadelphia: Benjamins.

Hall, Robert A. Jr. 1987. Bloomfield and semantics. In Hall. 155–60.

Hockett, Charles F., and Robert A. Hall Jr. 1987. A new Leonard Bloomfield bibliography. In Hall. 221–3.

Hoenigswald, Henry M. 1987. Bloomfield and historical linguistics. In Hall. 73–88.

Lehmann, Winfried P. 1987. Bloomfield as an Indo-Europeanist. In Hall. 163–72.

Moulton, William G. 1987. Bloomfield as a dialectologist. In Hall. 139–54.

Wolff, John U. 1987. Bloomfield as an Austronesianist. In Hall. 173–8.

RENATE BORN

Boas, Franz

Central to Franz Boas' anthropology, with its subfields of prehistoric archeology, physical anthropology, ethnology, and linguistics, was the idea of culture as a symbolic system or construct of the mind. Accordingly, Boas focused on ethnology and linguistics while he paid less attention to prehistoric archeology and physical anthropology. Language had a special place in Boas' work because it represented the prime example of a human symbolic system and was the best means for classifying the indigenous cultures of North America. Boas was primarily interested in the Native languages of the American Northwest. While he was concerned with the genetic classification of native languages, the main thrust of his work was directed toward investigating categorization in languages and establishing a typology of the inner forms or systems of unconscious grammatical categories of Native American languages.

His early fieldwork for the British Association for the Advancement of Science under the supervision of Horatio Hale (1817–1896) led to grammatical sketches, vocabularies, and linguistic maps of the languages of British Columbia. Boas' lexical and grammatical data of Nootka and Kwakiutl allowed him to unite the two languages as the Wakashan family (1890, 1891a, b). Based on grammatical evidence only, he also made the controversial proposal that Tlingit and Haida were related (1894). Boas' lifelong preoccupation with the languages of this region resulted in grammars of Tsimshian, Chinook (1911a), Tlingit (1917a), and Kwakiutl (1911a, 1947), and collections of native language texts. Boas further noted that some linguistic characteristics were distributed geographically among these languages, an observation that would eventually lead to the establishment of areal-typological studies.

Boas' ideas on linguistic categorization were rooted in nineteenth century German acoustics, psychophysics, and psychology. His exposure to so-called 'alternating sounds', speech sounds that were seemingly vague, during his fieldwork in British Columbia was the reason for a seminal paper, arguing that the vagueness of these sounds was a result of different categorizations of speech sounds in distinct languages and of their varying apperception by speakers of other languages (1889). Eventually, Boas was to argue that linguistic categorization, in general, was the outcome of grouping experiences into classes according to their similarity; however, the criteria underlying the similarity and class membership of experiences were not the same in all languages. In English, for example, the existence of the words *to grip*, *to kick*, *to pound*, and *to bite* indicated that the different activities they designated were classified as separate concepts. By contrast, in Dakota the form *xtaka* 'to grip', which was present in the terms *naxta'ka* 'to kick', *boxta'ka* 'to pound', and *yaxta'ka* 'to bite' suggested that the same activities

were classed under one concept only (1911a). Because linguistic categories did not rise to the level of consciousness, they provided a unique view on the mental processes of abstraction and association.

Influenced by the work of Heyman Steinthal (1823–1899), Boas suggested that the inner forms of languages molded their speakers' thoughts. The unconscious nature of these obligatory categories resulted in their projection onto reality and their objectification. Late in life Boas conceded, however, that although thought was 'directed in various channels' by grammatical categories, their influence was not to be overrated (1942:183). His linguistic relativism was further counterbalanced by the observation that all languages possessed the pronouns *I*, *thou*, and *he* and by his recognition of universal functional relations, such as subject–predicate, noun–attribute, and verb–adverb (1901, 1838). Boas' views suggest that he considered the universal occurrence of these pronouns and functional relations as examples of what Adolf Bastian (1826–1905) had termed *Elementargedanken*, elementary ideas that necessarily occurred everywhere because of the constitution of the human mind. Nevertheless, these ideas were not identical in different contexts but were shaped by different psychological factors and environmental circumstances. One important task of anthropology was the identification of these ideas.

Like Steinthal, Boas therefore objected to the use of grammatical categories derived from Indo-European languages in describing Native American languages. Such a practice distorted the nature of these languages and covered up their 'characteristic psychological categories' (1917b:5–6). Only analytical grammars, which treated languages on their own terms, were acceptable; only they would provide access to the inner form and 'psychological groupings' of a language (1911a:81). These grammars formed the basis for tracing the 'the history of the mental development of various branches of mankind' and for identifying grammatical concepts shared by the languages of the world, albeit in modified form (1911a:71).

The *Handbook of American Indian Languages* (1911–1941) was a monument to Boas' ideas and scholarship. He not only edited the four volumes comprising the *Handbook* but also contributed extensively to the first volume (1911a). Boas wrote the introduction and the grammars of Tsimshian, Kwakiutl, and Chinook and coauthored the grammar of Siouan (Boas and Swanton 1911). His introduction outlined his ideas on language and discussed its place within his anthropological framework. He showed that there was no necessary correlation between culture, language, and race, but that the three were independent of each other. Boas highlighted at length the process of cate-

gorization responsible for producing distinct phonetic, grammatical, and lexical systems characterizing different languages. The *Handbook* grammars were to illustrate the 'psychological principles' underlying Native American languages and to form the basis for establishing a typology of inner forms (1911:v). Boas provided an overview of the grammatical characteristics of Native American Languages and the principles of language classification and reduced John W. Powell's (1834–1902) 58 Native American language families to 55.

Early in his career, Boas (1894) had given more weight to grammatical similarities in determining relationships between languages. When he wrote his introduction to the *Handbook* (1911a), he had moved to a more balanced position that equally stressed phonetic, lexical, and grammatical similarities in classifying languages into families. Simultaneously, Boas increasingly emphasized an alternative to the common origin of seemingly related languages: multiple origins resulting from borrowing of linguistic elements and language mixture (1911a, 1917, 1920, 1929). Boas' late papers on the influence of Spanish on Nahuatl (1930) and of English on Kwakiutl (1932) documented instances of lexical and morphological intermixture between languages.

Biography

Franz Boas was born in Minden, Westfalia, Germany on July 9, 1858. He received a doctorate (1881) from the University of Kiel, Germany for his dissertation on the perception of the color of water. He undertook an expedition to Baffin Island in 1883–1884; habilitation (1886): Berlin University, Germany with a thesis on the geographical results of his Baffin Island trip. He moved to the United States in 1886, and became a US citizen in 1891. He was geographical editor of *Science* in 1887–1888. He carried out intermittent fieldwork for the British Association for the Advancement of Science (BAAS) in British Columbia in 1888–1894. He held a docentship in physical anthropology, Clark University in 1889–1892. He was assistant to F. W. Putnam, Head of the Department of Ethnology and Archaeology at the World Columbian Expedition in 1893–1894. He was assistant curator, American Museum of Natural History in 1895–1905. In the field, he was a member of the Jesup Expedition in 1897, and 1900. He began lecturing at Columbia University in 1896, and received full Professorship in 1899–1936. He was a member of the editorial board and publication committee of the *American Anthropologist*, in 1899–1920; editor of the *Publications of the Jesup North Pacific Expedition*, in 1900–1930; editor of *Journal of American Folklore* in

1908–1925; founder and editor of the *International Journal of American Linguistics* in 1917–1939; editor of the *Columbia University Contributions to Anthropology* in 1913–1936; and editor of the *Handbook of American Indian Languages* in 1911–1941. He was Vice-President of Section H. of the American Association for the Advancement of Science (AAAS) in 1894; President of the American Folklore Society in 1900–1901; Honorary Philologist, Bureau of American Ethnology in 1901–1920; President of the American Anthropological Association in 1907 and 1908; Vice-President and Chairman of Section H. of the AAAS in 1909; President of the 23rd International Congress of Americanists in 1928; President of the Linguistic Society of America (LSA) in 1928–1929; Chairman of the Committee of the American Council of Learned Societies on Research in American Indian Languages; Honorary President of the 24th International Congress of Americanists in 1930; and President of the American Association for the Advancement of Science in 1931. Franz Boas died in New York, NY, USA on December 21, 1942.

References

Boaz, Franz. 1911a. Handbook of American Indian languages, Part I, Bureau of American Ethnology, Bulletin 40, Washington, DC: Government Printing Office.

Boaz, Franz. 1889. On alternating sounds. American Anthropologist 2.

———.1890. First general report on the Indians of British Columbia. Report of the 59th Meeting of the British Association for the Advancement of Science, 1889. London: Murray.

———.1891a. Second general report on the Indians of British Columbia. Report of the 60th Meeting of the British Association for the Advancement of Science, 1890. London: Murray.

———.1891b. Vocabularies of the Tlingit, Haida, and Tsimshian languages. Proceedings of the American Philosophical Society 29.

———.1894. Classification of the languages of the North Pacific Coast. Memoirs of the International Congress of Anthropology, Chicago 1893, ed. by C.S.Wake. Chicago: Schulte.

———.1901. The mind of primitive man. Journal of American Folklore 14.

———.1909. Notes on the Iroquois language. Putnam Anniversary Volume. New York: G. E. Stechert.

———.(ed.) 1911a, 1922, 1941. Handbook of American Indian languages, Parts 1–2, Bureau of American Ethnology, Bulletin 40, Washington, DC: Government Printing Office; Part 3, Glückstadt–Hamburg–New York: J.J. Augustin, 1933–1938; Part 4, New York: J.J. Augustin.

———.(ed.) 1911a. Introduction. In Boaz.

———.(ed.) 1911a. Tsimshian. In Boaz.

———.(ed.) 1911a. Kwakiutl. In Boaz.

———.(ed.) 1911a. Chinook. In Boaz.

———.1911b. The mind of primitive man. New York: MacMillan.

———.1917a. Grammatical notes on the language of the Tlingit Indians. The University Museum Anthropological Publications 8.

———.1917b. Introductory. International Journal of American Linguistics 1.

———.1920. The classification of American languages. American Anthropologist n.s. 22.

———.1929. Classification of American Indian languages. Language 5.

———.1930. Spanish elements in Nahuatl, Todd Memorial Volumes Philological Studies, Vol. 1, ed. by John D. Fitz-Gerald and Pauline Taylor. New York: Columbia University Press.

———.1932. Notes on some recent changes in the Kwakiutl language. International Journal of American Linguistics 7.

———.(ed.) 1938. Language. General Anthropology. Boston: D.C. Heath.

———.1942. Language and culture. Studies in the history of culture: the disciplines of the humanities, ed. by P. Long. Freeport, NY: Books for Libraries Press.

———.1947. Kwakiutl grammar. Transactions of the American Philosophical Society 37, ed. by H. Boas Yampolsky, Z.S. Harris and John Swanton; 1911a. Siouan. In Boaz.

Cole, Douglas. 1999. Franz Boas: the early years 1858–1906. Vancouver/Toronto: Douglas and Mcintyre; Seattle and London: University of Washington Press.

Darnell, Regna. 1998. And along came Boas: continuity and revolution in Americanist anthropology. Amsterdam and Philadelphia: John Benjamins.

Goldschmidt, W. (ed.) 1959. The anthropology of Franz Boas. Essays on the centennial of his birth. Memoirs of the American Anthropological Association 89.

Gruber, J. W. 1967. Horatio Hale and the development of American anthropology. Proceedings of the American Philosophical Society 111.

Mackert, Michael. 1993. The roots of Franz Boas view of linguistic categories as a window to the human mind. Historiographia Linguistica 20.

Mackert, Michael. 1994. Franz Boas' theory of phonetics. Historiographia Linguistica 21.

Mackert, Michael. 1995. The Hale–Boas debate on the origin and classification of languages. History of Linguistics 1993, ed. by Kurt R. Jankowsky. Amsterdam and Philadelphia: John Benjamins.

Mackert, Michael. 1999. Franz Boas' early Northwest coast alphabet. Historiographia Linguistica 26.

Rohner, Ronald. 1966. Franz Boas ethnographer on the Northwest Coast. Pioneers of American Anthropology: The Uses of Biography, ed. by June Helm. Seattle and London: University of Washington Press.

Stocking, George W. 1974. The Boas plan for the study of American Indian languages. Studies in the History of Linguistics, ed. by Dell Hymes. Bloomington: Indiana University Press.

Stocking, George W. (ed.) 1996. History of anthropology. Vol. 8: Volksgeist as method and ethic: essays on Boasian ethnography and the German anthropological tradition. Madison: University of Wisconsin Press.

Williams, Vernon J. 1996. Rethinking race: Franz Boas and his contemporaries. Lexington: University of Kentucky Press.

MICHAEL MACKERT

See also **Chinook Jargon**

Body Language

See **Paralanguage**

Bopp, Franz

Franz Bopp was one of the founders of nineteenth-century comparative philology. He is often given credit for establishing historical linguistics as a serious area of scientific study. He was in a group of pioneering German scholars (such as the brothers August Wilhelm and Friedrich von Schlegel and Wilhelm von Humboldt) that started their study of Sanskrit and Indo-European at the University of Paris and, later on, made it a German monopoly.

Bopp's chief activity centered on the study of the languages of the East. His knowledge of the literature written in Sanskrit and the language itself was truly expert. Bopp was keenly interested in the great Indian epics. He published several episodes from the MahâBhârata, both in Sanskrit and in translated form. He also compiled a Sanskrit and Latin glossary. Many European researchers (e.g. William Carey or Henry Thomas Colebrooke) had already proposed their own grammars of Sanskrit, based on the Indian tradition that stretched back to Panini. Such grammars were obscure and unhelpful from the point of view of scholars who were used to the Western model of paradigms. In his very influential Sanskrit grammar, Bopp abandoned much of the original apparatus of the Indian system of grammatical description. Instead, he adopted the European framework that was grounded in the traditional analyses of Greek and Latin.

Bopp's Sanskrit studies were especially important as the requisite stimulus for great developments in the field of comparative grammar. Bopp traced many similarities between Sanskrit and its Western cognate languages, both in terms of lexical items (roots) and on the syntactic level. He denied the importance of surface differences between grammars, claiming that they do not matter as long as it is possible to reconstruct the way languages have developed from the common source. Therefore, one of his prime objectives was to investigate the oldest accessible form of a given language.

Bopp wrote many monographs on individual languages (e.g. Celtic, Old Prussian, and Albanian) and on languages outside the Indo-European family (e.g. a paper on the relationship between Indo-European and Malayo-Polynesian). His investigation and systematic comparison of a number of languages was a tremendous refinement of descriptive techniques used by linguists at that time.

In his most important work, *Vergleichende Grammatik des Sanskrit, Zend, Griechischen, Lateinischen, Litauischen, Altslavischen, Gothischen und Deutschen* (1833–1852 ; *A Comparative Grammar of the Sanskrit, Zend, Greek, Latin, Lithuanian, Gothic, German, and Slavonic Languages*), Bopp discussed and established relationships between the Indo-European languages. In principle, a resemblance between related languages could be explained in two ways: either by claiming that one of them is derived from the other or by assuming that both are derived from another language. Contrary to many of his contemporaries, Bopp did not regard languages such as Greek or Latin as derived from Sanskrit. He describes Indo-European languages as variations of one original language. However, Sanskrit has preserved more features of the common source than other cognate languages.

Since the publication of his first major work *Über das Conjugationssystem der Sanskritsprache in Vergleichung mit jenem der griechischen, lateinischen, persischen und germanischen Sprache* (1816; On the System of Conjugation in Sanskrit, compared with the Greek, Latin, Persian, and German Languages), Bopp's main interest was focused on detecting the common origin of Indo-European inflections. Primarily, he applied his analysis only to the verb. However, while in London, he published a very influential essay in the Annals of

Oriental Literature (entitled *Analytical comparison of the Sanskrit, Greek, Latin and Teutonic languages, shewing the original identity of their grammatical structure*), in which he included analysis of nonverbal inflection. This essay received considerable public attention, which resulted in Bopp's appointment to the newly established chair of Oriental Literature and General Philology at the University of Berlin in 1821 (an additional factor was Bopp's friendship with Wilhelm von Humboldt, who was at that time the Prussian Minister of Worship and Public Instruction).

Rejecting Friedrich Schlegel's ideas, Bopp advocated an agglutinative theory of the origin of morphological structure. He assumed that grammatical modifications can be expressed only by external additions to the root (i.e. by means of agglutination of separate lexical items which become suffixes). According to Bopp, what Schlegel understood as 'internal' inflexion (a change of root vowels or reduplication) is an example of redundancy in Indo-European—it does not indicate any modification of meaning. Phenomena such as apophony cannot be semantically relevant since they are not separate elements. Bopp analyzed roots as the primitive (monosyllabic) elements of words that can be identified through examination of forms that contain them as a common base. There are two types of roots: verbal (predicative) and pronominal (indicative). A root must be considered to be an abstract entity, which means it does not have to be itself equivalent to any existing word.

Bopp distinguished those aspects of languages that can be formally described from those that are arbitrary. The former comprise what he called 'mechanical' (i.e. phonetic) laws and the origin of the forms indicating grammatical relationships (morphology). On the other hand, in his *Vergleichende Grammatik*, Bopp refers to the relationship between form and meaning as the 'secret of roots'—i.e. he notices that it is impossible to say why a given root means what it means, e.g. why the Indo-European root STHA/STA means 'stand' and not 'go'.

Nineteenth century philologists tended to use biological terminology in describing the way that languages work. Being influenced by contemporary developments of comparative anatomy, they viewed languages as changing and organic. It could be said that Bopp was no exception. He also tried to dissect and classify languages like living organisms. However, he examined them as static entities; the historical process of change did not interest him as much as the structure that resulted from it.

Bopp devoted his life to laborious, quiet study in libraries, remaining aloof from any political or social events. He was a zealous researcher, very cautious and careful, constantly maturing and correcting his works (many of them were reissued with important improvements). He wanted his conclusions to be as objective and nonspeculative as possible. Therefore, he was (and is) sometimes called 'the factual Bopp'.

Biography

Franz Bopp was born in Mainz, Germany on September 14, 1791. He attended lectures in law, logic, esthetics, history, and philosophy at the Aschaffenburg Ecole du Droit, Bavaria, Germany. He moved to Paris, France, to study Sanskrit (together with August Wilhelm von Schlegel and Wilhelm von Humboldt, tutored by Antoine Léonard de Chézy) in 1812–1817, resided in London, England in 1817, and continued his studies in Göttingen, Germany in 1818–1821. He was Extraordinary Professor of Oriental Literature and General Philology, University of Berlin, Germany in 1821 and Ordinary Professor of Oriental Literature and General Philology, University of Berlin, Germany in 1825 (he held this post till death). He was also a member of the Royal Prussian Academy in 1822 and honorary member of the Philological Society. Bopp died in Berlin on October 23, 1867.

References

Bopp, Franz. 1816. Über das Conjugationssystem der Sanskritsprache in Vergleichung mit jenem der griechischen, lateinischen, persischen und germanischen Sprache.

———. 1820. Analytical comparison of the Sanskrit, Greek, Latin and Teutonic languages, showing the original identity of their grammatical structure. Annals of Oriental Literature 1; reprinted as Analytical comparison of the Sanskrit, Greek, Latin and Teutonic languages, showing the original identity of their grammatical structure. Newly edited, together with a bio-bibliographical account of Bopp by J. D. Guigniaut, an introduction to Analytical comparison by F. Techmer and a letter by W. von Humboldt, 1974.

———. 1827. Ausführliches Lehrgebäude der Sanskrit-Sprache.

———. 1829–1832. Glossarium sanscritum; new edition Glossarium sanscritum, in quo omnes radices et vocabula usitatissima explicantur et cum vocabulis graeis, 1840–1847.

———. 1833–1852. Vergleichende Grammatik des Sanskrit, Zend, Griechischen, Lateinischen, Litauischen, Altslavischen, Gothischen und Deutschen; as A comparative grammar of the Sanskrit, Zend, Greek, Latin, Lithuanian, Gothic, German, and Slavonic languages, transl. by Edward Backhouse Eastwick, 1856; revised edition Vergleichende Grammatik des Sanskrit, Zend, Armenischen, Griechischen, Lateinischen, Litauischen, Altslavischen, Gothischen und Deutschen, 1857–1861.

———. 1834. Kritische Grammatik der Sanskrita-Sprache in kürzerer Fassung.

———. 1841. Über die Verwandtschaft der malayisch-polynesischen Sprachen mit den indisch-europäischen.

Lefmann, Salomon. 1891–1895. Franz Bopp, sein Leben und seine Wissenschaft, 2 vols. Berlin: G. Reimer.

Lehmann, Winfred P. 1991. Franz Bopp's use of typology. Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 44(3).

Lehmann, Winfred P. 1996. Theoretical bases of Indo-European linguistics. London and New York: Routledge.

Leonhard, Tafel. 1861. A review of some points in Bopp's comparative grammar. Andover, MA: W.F. Draper.

Leskien, August. 1876. Bopp. Allgemeine deutsche Biographie 3. In Seboek.

Martineau, Russell. 1867. Obituary of Franz Bopp. Transactions of the Philological Society 12, reprinted as Franz Bopp. In Seboek.

Seboek, Thomas A. (ed.) 1966. Portriats of linguists: a biographical source book for the history of Western linguists, 1746–1963, Vol. 1. From Sir William Jones to Karl Bruymann. Bloomington: Indiana University Press.

Sternemann, Reinhard. 1984. Franz Bopp und die vergleichende indoeuropäische Sprachwissenschaft: Beobachtungen zum Boppschen Sprachvergleich aus Anlass irriger Interpretationen in der linguistischen Literatur. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Sternemann, Reinhard (ed.) 1994. Bopp-Symposium 1992 der Humboldt-Universität zu Berlin: Akten der Konferenz vom 24.3-26.3.1992 aus Anlass von Franz Bopps zweihundertjährigem Geburtstag am 14.9.1991. Heidelberg: Universitätverlag C. Winter.

Verburg, Pieter A. 1950. The background to the linguistic conceptions of Franz Bopp. Lingua 2.

PAWEŁ RUTKOWSKI

See also **Indo-European 1: Overview; Panini; Philology; Sanskrit**

Brain Organization and Auditory Pathway

Three major functions of the brain are the integration of sensory input, behavior in general, and coordination of body functions. The development of the brain depends upon both the multicellular structure of creatures and the necessity to control complex behavior of highly developed organisms. Before multicellular organisms evolved, a unicellular organism (Protozoa) already represented an omnipotent life form, which performed e.g. movement, ingestion, and orientation with merely one cell. After the evolutionary emergence of multicellular organisms (Metazoa), certain cells of these organisms adapted functionally and became e.g. secretory, supporting, sensory, or muscle cells. As a result, during this developmental stage an increasing number of specialized body cells had to coordinate with each other; hence, a communication network became necessary.

Communication took place by an additional specialized cell type—the nerve cell. Starting from the stage of simple nerve nets (e.g. in jellyfish), the evolution of nervous systems in animals led to small groups of nerve cells (ganglia), peripheral nervous systems, and central nervous systems (brains). The human central nervous system represents the most complex integration/coordination system: it comprises about 100 billion (10^{11}) nerve cells, each connected, on average, to about 1,000 others. During the nine months of pregnancy, the brain of a human fetus has to grow on average by 250,000 new nerve cells per minute in order to reach its final dimension (on average 1,400 gram). However, after birth the brain mainly controls the basic body functions. The adult brain is not only involved in the coordination of organs, body functions, or parts but also in integrating sensory input, learning, adaptive behavior, and cognition. To reach its full functional state, the newborn's brain has to 'mature' for some additional months, during which period the neural networks for higher sensory and basic cognitive processes have to establish themselves. Beginning already in the uterus, where the unborn child learns parts of the mother's phoneme set, the child's brain needs about two additional years to produce language after birth. This shows the spectrum of the different brain capabilities from controlling body functions to cognitive processes. From an evolutionary point of view, the brain consists of three overlaying brain parts: (1) the approximately 250-million-year-old 'reptile brain', performing coordination of body functions and simple sensory integration, (2) the 'older mammalian brain' performing e.g. complex integration of sensory input and emotions, and (3) the 'only' 50-million-year-old 'young mammalian brain' (neocortex), the neuronal substrate for higher cognitive processes. Even though these brain parts form a functional unit and act as a single organ, these three phylogenetic stages can be assigned to different levels of cognitive performance. Language corresponds to the most developed cognitive process and is only developed in humans.

Due to its importance, the human central nervous system, which consists of the brain and the spinal cord, is protected by three membranous coverings (meninges) and by bony capsules (cranium, backbone). Three main parts of the brain are shown in a side view (see Figure 1): the brain stem (B), the cerebellum (C), and the left cerebral hemisphere (H). The two cerebral hemispheres represent 85% of the brain's weight and are surrounded by the 1.3 to 4.5 mm thick cerebral cortex. The cortex
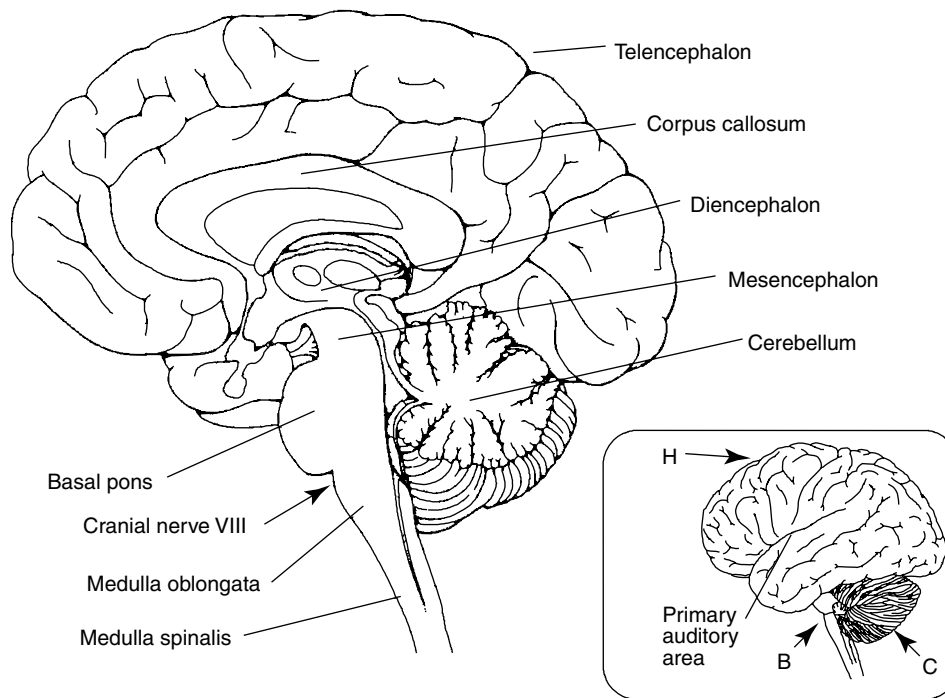
Figure 1. Brain Organization and Auditory Pathway.

consists of specially organized nerve and glial cells and is compactly folded, which causes its surface enlargement, although the volume remains unchanged. The cortex shows an abundant structure of ridges (gyri) and valleys (sulci) and represents the main neuronal structure for higher cognitive processes. Each hemisphere is divided into four lobes: the frontal, the temporal, the parietal, and the occipital lobe. Anatomically, the brain is divided into five parts, which are shown in Figure 1: the myelencephalon (medulla oblongata), the metencephalon (pons, cerebellum), the mesencephalon, the diencephalon, and the telencephalon (cerebral hemispheres). A major structure for transferring information between both hemispheres is the corpus callosum, a broad nerve tract that also belongs to the telencephalon.

One of the brain's major function lies in the integration of sensory information from the auditory channel. For primates, the auditory input can be seen as the most important input from the environment. Like every other audible sound, language is perceived by the ears and transferred to certain cortical regions across the so-called auditory pathway. The auditory pathway lead from the nerve fibers of the cochlear branch of the VIII cranial nerve (vestibulocochlear nerve) via the primary auditory nuclei of the medulla, to several neural centers of the brain stem, up to the primary and the secondary auditory cortex. Most of the structures analyzing the auditory signal are located in the brain stem, which comprises the medulla, the pons, and the mesencephalon. Since the internal ear lies close to the brain stem, the auditory nerve is only about 5 mm away from

the medulla, near the inferior border of the pons, into which the nerve inserts (see Figure 1).

At first, the sound signal is gathered and filtered by the external ear. After mechanical amplification in the middle ear, the signal is transduced into an electrical signal by sensory cells (organ of Corti) in the inner ear. The organ of Corti of each ear contains about 3,500 inner hair cells and about 12,000 outer hair cells, which are innervated by 30–40,000 fibers of the cochlear branch of the VIII cranial nerve (auditory nerve). About 90% of the fibers are connected to inner hair cells, and these mainly contribute to the information on sound signals for subsequent analysis in the brain. The analysis of the sound signal starts already in the cochlea, e.g. by separating the signal into the frequency domain (frequency dispersion), which will not be discussed here in detail. The major part of the analysis (feature detection) is carried out during the next stations of the projection to the auditory cortices. The detailed function of the auditory cortex is still not understood—however, it is the final place of sound recognition, especially for speech analysis (e.g. phoneme detection). Starting at the cochlea, at each level of the auditory pathway, the neurons are arranged in a well-ordered map that reflects the frequencies of the stimuli (tonotopic organization). Consequently, neurons responsive to low and high frequencies are located at a maximum distance from each other.

While each hemisphere receives sensory input and controls motor movement of the opposite (contralateral) side of the body, the information of each ear goes

to both hemispheres, whereby the majority crosses to the contralateral side. First, all afferent fibers of the auditory nerve terminate in the ipsilateral cochlear nucleus, whereas during projection a major part of the afferences change to the contralateral side of the brain (indicated by line thickness in Figure 2). Due to this crossing of fibers at several stations and the presence of decreasing fibers and parallel processing, the audi-

tory system exhibits increased stability against damage to lower neural structures. The next station of the pathway, the superior olivary complex, already receives bilateral input from both cochlear nuclei. The main processing stations in the superior olivary complex are the medial superior olive, the lateral superior olive, and the medial nucleus of the trapezoid body. The next major station of the auditory pathway is the
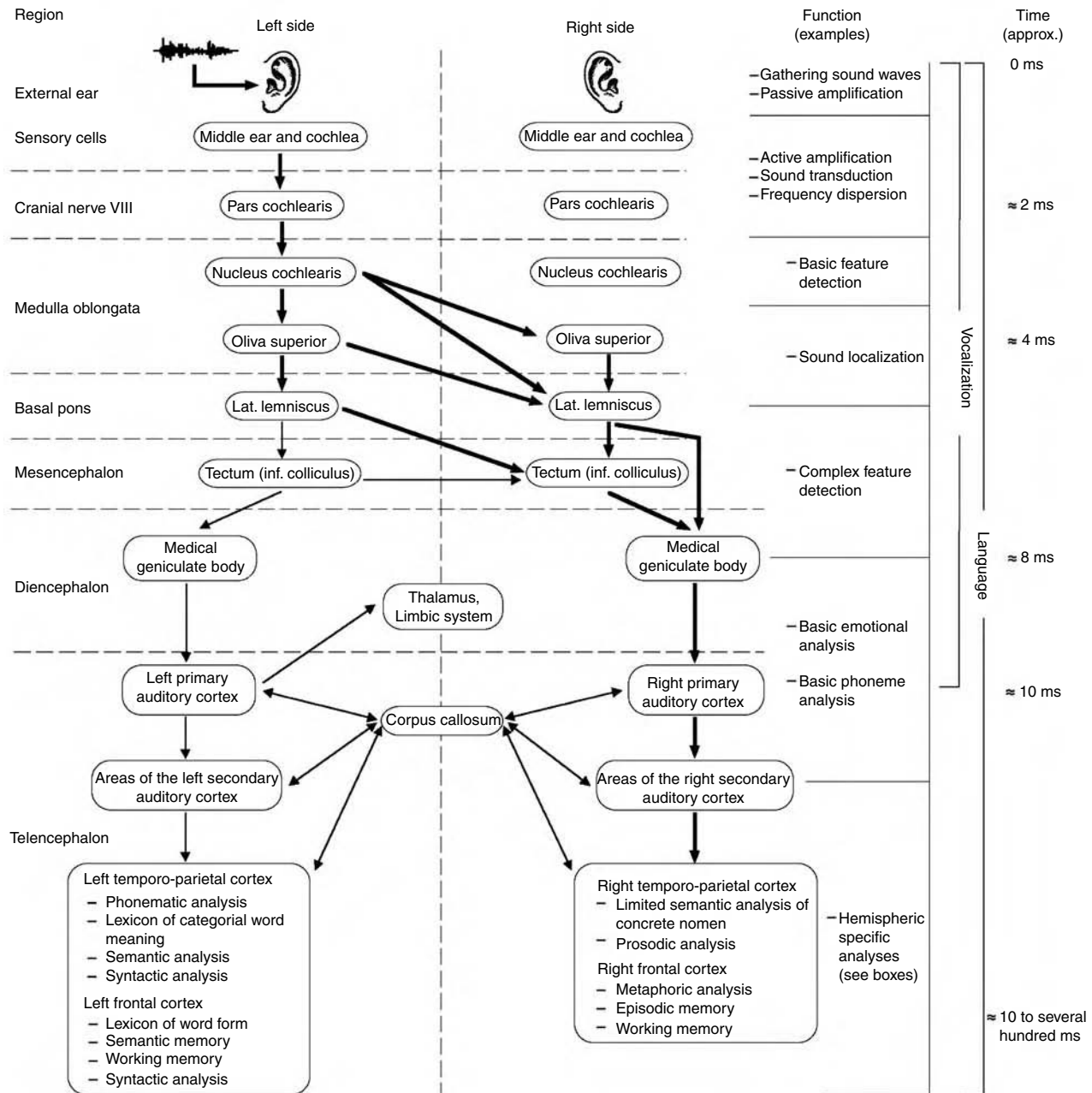
Simplified illustration of the auditory pathway of the left ear and some examples of known processing stages for spoken language comprehension. The identical, mirrored pathway of the right ear is not shown.

Figure 2. Brain Organization and Auditory Pathway.

lateral lemniscus, which is divided into a ventral and a dorsal nucleus. The following two major stations are the inferior colliculus and the medial geniculate body. The primary auditory cortex, which is identical to Brodmans area 41, comprises only a small area of the superior surface of the superior temporal gyrus (Heschl's gyrus) and is the principal target of all ascending auditory pathways (see Figure 2).

In most people, the left hemisphere plays a more important role than the right hemisphere with regard to language processing. Due to the multiple crossing of the signals of each ear, damage in the lower auditory nucleus of one hemisphere does not lead to a unilateral, complete loss of acoustic processing, but results in a bilateral decreased hearing ability. However, substantial left hemispheric brain lesions of language-related cortical areas do not reduce the hearing ability, but only lead to language-specific impairments (see chapters 'Lateralization and Handedness' and 'Aphasia').

References

Aitkin, Lindsay. 1990. The auditory cortex: structural and functional bases of auditory perception. London: Chapman & Hall.

Brown, M. Christian. 1999. Audition. Fundamental neuroscience, ed. by Michael J. Zigmond, Floyd E. Bloom, Story C. Landis, James L. Roberts and Larry R. Squire. San Diego: Academic Press.

Dearmond, Stephen J., Madeline M. Fusco, and Maynard M. Dewey. 1989. Structure of the human brain: a photographic atlas, 3rd edition. Oxford: Oxford University Press.

Durrant, John D., and Jean H. Lovrinic. 1995. Bases of hearing science, 3rd edition. Baltimore: Williams & Wilkins.

Gazzaniga, Michael S. (ed.) 2000. The new cognitive neurosciences. Cambridge, MA: MIT Press.

Gummer, Anthony W., and Hans-Peter Zenner. 1996. Central processing of auditory information. Comprehensive human physiology, Vol. 1, ed. by Rainer Greger and Uwe Windhorst. Berlin: Springer.

Kandel, Eric R., James H. Schwartz, and Thomas M. Jessel. 1991. Principles of neural science, 3rd edition. New York: Elsevier.

Moore, Brain C.J. (ed.) 1995. Hearing 2nd edition. San Diego: Academic Press.

Nolte, John. 1999. The human brain: an introduction to its functional anatomy, 4th edition. St. Louis: Mosby.

Purves, Dale, George J. Augustine, David Fitzpatrick, Lawrence C. Katz, Anthony-Samuel LaMantia, and James O. McNamara (eds.) 1997. Neuroscience. Sunderland: Sinauer Associates.

Webster, William R., and Laurence J. Garey. 1990. Auditory system. The human nervous system, ed. by George Paxinos. San Diego: Academic Press.

Zemlin, Willard R. 1998. Speech and hearing science: anatomy and physiology, 4th edition. Boston: Allyn & Bacon.

HORST M. MÜLLER

Bresnan, Joan

Few students enter Joan Bresnan's classes in syntactic theory and leave without being fascinated about the subject matter of syntax. She is one of the most influential teachers and researchers in the field of syntax in the late twentieth and early twenty-first centuries. Joan Bresnan entered graduate school at the time when transformational approaches to syntax, as developed mainly by her teacher at MIT, Noam Chomsky, constituted the dominant paradigm. Transformational approaches to the analysis of syntactic representation often take one level of representation as the basic point of departure and attempts are then made to derive other levels of representation from this basic level through a series of transformational operations, summarized as move alpha. The most classical example is the passive transformation whereby a passive sentence in a language is derived by applying a passive rule on an active sentence. However, confronted with a lot more natural language data from languages with structures very different from those of the known European languages like English and French, it became clear that other approaches to the analysis of syntax were necessary to address the problems of the transformational approach.

It was at this point that Joan Bresnan drew attention to the role of the lexicon in syntactic analysis. The lexicon in early days of transformational grammar played only a derivative role in syntactic analysis as it was regarded as having too many lexical idiosyncracies in a framework that stressed the role of transformational rules. With evidence from various languages, such as Chichewa and other Bantu languages, native Australian languages, and native American languages, Joan Bresnan demonstrated the important role that lexical information ought to play in syntactic theory. This led to her first major revolution in syntatic theory: the development of the idea of Lexical-Functional Grammar (LFG), and then making it one of the most influential linguistic theories/frameworks of our time. Joan Bresnan developed LFG with Ronald Kaplan in the late 1970s and early 1980s. The term 'Lexical-Functional Grammar' first appeared in Bresnan's edited volume in 1982, *The mental representation of*

grammatical relations. It is a nontransformational and nonderivational approach to the representation of grammatical information. It is lexically oriented and assumes the basic idea that different dimensions of grammatical information can be represented as separate but parallel structures. Grammatical information is represented in LFG in three main structures: the Constituent Structure, where precedence and dominance relations are indicated, the Functional Structure, where grammatical functions such as SUBJECT and OBJECT are indicated, and the Argument Structure, where grammatical roles such as AGENT and THEME are expressed. These levels interact through a system of mapping or correspondence rules.

Joan Bresnan's second major revolutionary contribution to syntactic theory came with the rise of Optimality Theory (OT) in the early 1990s. In this theory of grammar, Universal Grammar is characterized as a set of constraint interactions, with the basic hypothesis that there is a set of universal constraints. Languages satisfy these constraints differently and differences in languages will result from this fact. This approach to grammar is potentially compatible with many theories of grammar that assume an input from which a set of candidate outputs are generated. It was at this point that Joan Bresnan came up with the idea of marrying the basic tenets of OT with those of LFG to form the notion of OT-LFG or what she termed 'Optimal Syntax'. In Optimal Syntax, the universal input is assumed to be modeled by an abstract set of functional structures representing morphosyntactic content in a language-independent format. The universal candidate set in Optimal Syntax should consist of pairs of constituent structures and their corresponding functional structures, which are then matched to the input functional structure through a set of correspondence principles and constraints that decide the optimal candidate from the set of candidates. Optimal Syntax or OT-LFG is one of the most exciting approaches to grammatical studies in the early part of the twenty-first century and this is being led by Joan Bresnan.

In addition to these major contributions in syntactic theory, Joan Bresnan, with her students and colleagues, has also made profound contributions to the description and analysis of particular languages and language groups, including Bantu languages such as Chichewa, and Australian aboriginal languages such as Warlpiri.

Biography

Joan Bresnan was born in Chicago, Illinois, on August 22, 1945. She received a B.A. degree in Philosophy at Reed College in 1966. She completed a Ph.D. on the theory of complementation in English under the supervision of Noam Chomsky at MIT in 1972. From 1972 to 1973, she was Assistant Professor of Linguistics and Philosophy at Stanford University. In 1973, she moved to the University of Massachusetts at Amherst and served as Assistant Professor from that time until 1975. Between 1979 and 1983, she was full Professor at MIT. Joan Bresnan moved back to Stanford University in 1983 as Professor of Linguistics and has remained there to date. Professor Bresnan held a Howard H. and Jessie T. Watkins University Professorship between 1992 and 1997, and in 2000 she was named Sadie Dernham Patek Professor in Humanities. Joan Bresnan has received many awards and fellowships, and is a member of many societies and boards. She was a Guggenheim Fellow from 1975 to 1976, a Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford University from 1982 to 1983, and a principal investigator of many projects financed by National Science Foundation Grants. She is a member of the editorial boards of many journals and publishing houses such as Cambridge University Press, Linguistic Inquiry, and the Journal of Japanese Linguistics. She is a past member of the Executive Committee of the International Lexical-Functional Grammar Association, and Past President of the Linguistic Society of America.

References

Aissen, Judith. 1999. Markedness and subject choice in optimality theory. Natural language & linguistic theory, Vol. 17, also in Optimality-theoretic syntax, ed. by Geraldine Legendre, Sten Vikner, and Jane Grimshaw. Cambridge, MA: MIT Press.

Bresnan, Joan. 1979. The theory of complementation in English syntax. New York, NY: Garland Publishing.

———.(ed.) 1982. The mental representation of grammatical relations. Cambridge, MA: MIT Press.

———.1997. Complex predicates, co-edited with Alex Alsina and Peter Sells. Stanford, CA: CSLI Publications.

———.2000. Optimal syntax. Optimality theory: phonology, syntax, and acquisition, ed. by Joost Dekkers, Frank van der Leeuw and Jeroen van de Weijer. Oxford: Oxford University Press.

———.2001. Explaining morphosyntactic competition. The handbook of contemporary syntactic theory, ed. by Mark Baltin and Chris Collins. Malden, MA: Blackwell Publishers.

———.2001. Lexical-functional syntax. Malden, MA: Blackwell Publishers.

Kaplan, Ronald M. 1989. The formal architecture of lexical-functional grammar. Proceedings of ROCLING II, ed. by Chu-Ren Huang and Keh-Jiann Chen, also in Journal of Information Science and Engineering 5(4). 305–22.

Kaplan, Ronald M. 1995. The formal architecture of lexical-functional grammar. ed. by Dalrymple, et al., 7–27. Stanford, CA: CSLI Publications.

Kaplan, Ronald M., and Joan Bresnan. 1982. Lexical-functional grammar: a formal system for grammatical representation.

The mental representation of grammatical relations, ed. by Joan Bresnan. Cambridge, MA: MIT Press.

Sadler, Louisa. 1996. New developments in LFG. Concise encyclopedia of syntactic theories, ed. by Keith Brown and Jim Miller. Oxford: Elsevier Science.

Wescoat, Michael, and Annie Zaenen. 1991. Lexical functional grammar. Linguistic theory and grammatical description, ed. by Flip G. Droste and John E. Joseph, Current issues in linguistic theory, Vol. 75. Amsterdam: Benjamins.

ADAMS BODOMO

See also **Chomsky, Noam; Grammar, Theories; Lexicon: Overview; Universal Grammar**

British Sign Language

British Sign Language (BSL) is the language of the British Deaf community, with its associated social and cultural organizations. There are no accurate figures of the number of users of BSL, but estimates range between 30,000 and 60,000 deaf signers. Hearing children of signing deaf parents may also be included. The number of signers who are not members of the community has increased enormously over the past 15 years, through the widespread uptake of further education classes in BSL. In addition to signers in Britain, BSL is also used by some signers in Ireland, and it is taught as a foreign language to deaf school children in Norway and Russia.

Historical Records of BSL

BSL has no written form. Consequently, historical records of the language are rare. Of the few early recorded references to sign language use in Britain, most are in the form of descriptions in English by hearing writers. We do know that deaf people have been using some form of signing in Britain since at least the sixteenth century. A description of the signs used by a deaf man to make his marriage vows in 1575 is recorded in the parish records of St. Martin's in Leicester. From the mid-seventeeth century onwards, books such as those by Bulwer (1644, 1648) were published, which included drawings of signs and lexical and grammatical descriptions of BSL.

Also, in the mid-seventeenth century, the first manual alphabets appeared. These provide a means of reproducing the orthography of the written language using various configurations of the hands and fingers. Fingerspelling serves as a resource for lexical borrowing and new sign formation. Unlike most deaf communities, the BSL community uses a two-handed alphabet, the earliest version of which was published anonymously in 1698.

BSL was exported by signers to other countries, and was the predecessor of Auslan (Australian Sign Language) and New Zealand Sign Language. It has been suggested that these three languages should be considered as dialects of a single language. There is also evidence of the early influence of BSL on signing in South Africa, Hong Kong, and Newfoundland. Perhaps surprisingly, there are few known links between BSL and American Sign Language (ASL) and the languages are not mutually intelligible. BSL is also not genetically related to Irish Sign Language (ISL) despite their close geographical proximity.

Regional Dialects

There is considerable regional lexical variation in BSL, including core semantic areas such as color terms, days of the week, and numerals. Despite this variation, most signers are familiar with different dialects of BSL, and it appears likely that the use of signing on television over the past 20 years is gradually leading to the development of a national standard.

Grammar of BSL

BSL and other sign languages exhibit structural differences from spoken languages, as use of the visual–spatial modality provides a different range of structural possibilities, including visual imagery, movement in space, and multiple articulators.

In common with all sign languages, BSL signs often reflect visual properties of referents, such as their shape, movement, or how they are handled.

Whether or not a sign is visually motivated, all signs exhibit a conventionalized relationship between the form and the referent.

The placement and movement of signs within a linguistically determined area of space around the signer are a central feature of BSL, having considerable

grammatical functions. As in other sign languages, the movement of verbs between points in space can reflect grammatical relations between referents at the conceptual level, differentiating semantic roles and grammatical classes (subjects and agents, patients and objects of actions), without any implication of a spatially constrained relationship.

Signed language can also convey spatial relations more directly: sentences can be constructed topographically, with the spatial relationships between signs corresponding to actual relationships among the referents.

The availability of two hands, a head, and a face provides BSL and other sign languages with the possibility of using simultaneously articulated structures to place referents in space and represent their relative locations and movements. BSL can also directly represent the timing of two events relative to each other, by showing the two events on two different 'channels'. Sign languages differ in the extent to which they exploit such devices. In a comparison of sentences generated from the same picture materials in ISL and BSL, simultaneous signs appeared in 20% of ISL utterances and 80% of BSL utterances. Simultaneity can also be extended for poetic effect. In a BSL poem by Dorothy Miles, 'Trio', the poet represents herself, a dog, and a bird all dozing together after a good lunch. This is signed by using three articulators simultaneously, using one hand to refer to the dog, one hand to refer to the bird, and the head to refer to the poet.

BSL as a Minority Language

Although BSL is fully independent of English, both lexically and grammatically, English has influenced BSL, as might be expected when any powerful majority language surrounds a minority language. English provides loan signs to BSL through fingerspelling and loan translation.

Family terms, calendar vocabulary items, and signs for units of measurement of time and space in BSL are frequently of English origin. BSL also reflects the influence of English in its use of mouth patterns derived from spoken English ('mouthings'). These are used in a wide variety of ways and in conjunction with other mouth patterns unrelated to English ('mouth gestures'). The use of mouthings varies with age, and social and linguistic background, as well as situational variety. Although mouthings feature in all languages and function in similar ways, anecdotal evidence suggests that BSL uses many more mouthings than, for example, ISL or ASL.

While BSL is a minority language with respect to English and borrows from English, it also acts as a donor to other sign languages, for example, ISL.

Social Situation of BSL

The education of deaf children has had a great impact on BSL. The first schools for deaf children in Britain, using BSL as the language of instruction, were established in the mid-eighteenth century. Although BSL enjoyed widespread use in British deaf schools throughout the nineteenth century, there were always educationalists who believed that deaf children should use English rather than BSL. This philosophy gradually gained strength, and by the early twentieth century signing was banned in schools and children were punished for its use. Signing was reintroduced into deaf education from the 1970s onward, with the acceptance of bilingual (BSL and English) education. However, the closure of many deaf schools has reduced opportunities for deaf children to use sign language at an early age, and this has been viewed as a threat to the viability of BSL in the future, with the removal of children's access to the community of signers and the linguistic and cultural role model that the community provides to young deaf children.

Despite the changes in education policy, the past 20 years have witnessed a substantial improvement in the status of BSL. A weekly program in BSL has been broadcast by the BBC since 1981; a comprehensive dictionary of BSL was published in 1992; and the British government appears poised to offer some formal recognition to BSL as a minority language, and public interest in and acceptability of BSL has probably never been higher.

References

Boyes-Braem, Penny, and Rachel Sutton-Spence (eds.) 2001. The hands are the head of the mouth. Hamburg: Signum Press.

Brennan, Mary, Martin Colville, and Lillian Lawson. 1984. Words in hand: a structural analysis of the signs of BSL. Edinburgh: Moray House College of Education.

Brien, David (ed.) 1992. Dictionary of BSL/English. London: Faber & Faber.

Bulwer, John B. 1644. Chirologia: or the natural language of the hand. London: R. Whitaker.

Bulwer, John B. 1648. Philocophus: or the deafe and dumbe man's friend. London: Humphrey Moseley.

Johnston, Trevor. 1989. AUSLAN dictionary. A dictionary of the sign language of the Australian deaf community. Victroria: Australian Print Group.

Kennedy, Graeme, Robert Arnold, Pat Dugdale, and David Moskowitz. 1998. A dictionary of NZSL. Auckland: Auckland University Press.

Kyle, Jim, and Bencie Woll. 1985. Sign language: the study of deaf people and their language. Cambridge: Cambridge University Press.

Rachel Sutton-Spence, and Bencie Woll. 1999. The linguistics of British sign language: an introduction. Cambridge: Cambridge University Press.

BENCIE WOLL

See also **American Sign Language**

Brugmann, Karl

Best known as the leading member of the so-called 'Junggrammatiker' ('neogrammarians'), Karl Brugmann played a crucial role in laying the foundations of historical linguistics in the modern sense. In one of his most famous publications, the preface to the *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*, Brugmann and his co-editor, H. Osthoff, reacted to current ideas about language change, outlining a theory based on the regularity of sound laws, whose apparent irregularities could only be explained as either the outcome of other sound laws, or as the effect of analogy.

The importance of the concept of sound laws for the development of historical linguistics can hardly be overstated. Before the neogrammarians, phonological change was conceived as applicable in most cases, but no theoretical statement was made about its regularity, and the (apparent) exceptions were left unexplained. By giving regularity the status of a basic requisite, the neogrammarians opened the way to much more refined research in the field of sound change and language change in general, based on the idea that a phenomenon was not exhaustively described, until all possible exceptions or counterexamples had been satisfyingly explained.

Although the achievements of the neogrammarians are best described as the result of in-depth collaboration among a group of scholars who shared the same ideas, Brugmann certainly stood out of the group, with his 400 odd publications, covering various areas of linguistics and philology. The most famous remains his *Grundriß der vergleichenden Grammatik der indogermanischen Sprachen*, whose first volume appeared in 1886. The work, completed in 1901 with the addition of the second volume by Karl Brugmann and three volumes about syntax by Berthold Delbrück, is still the most authoritative reference book in the field.

By shifting emphasis from language classification and reconstruction to language change, the Neogrammarians broke with the tradition of Indo-European linguistics in two ways. In the first place, they brought forth a conception of linguistics as a historical science, thus closer to human sciences, as sociology and psychology, than to natural sciences, to which it was closer before. Furthermore, they provided linguistics with a methodology of its own, no longer borrowed from natural sciences, and not directly dependent on the methods of classical philology. This last point is of crucial importance for the emancipation of linguistics, and caused the most profound break with the scholars of the preceding generation. Brugmann's personal vicissitudes are an example of this break: at the inception of his scientific career, he produced two articles, *Nasalis sonans in der indogermanischen Grundsprache* and *Zur Geschichte der stammabstufenden Deklinationen*, which provoked a harsh reaction from his former teacher Georg Curtius, who, feeling criticized as methodologically inadequate, never wanted to acknowledge his pupil's achievements and, still years later, published a thorough, but by then outdated critique of the Neogrammarians.

Although his major interest was in the fields of phonology and morphology, Brugmann also published some pioneering work on syntax. He was deeply convinced of the importance of a global vision of utterances, prior to the study of forms in isolation, as shown by his statement, 'In fact, all speech is brought about in sentences' ('In der Tat geschiet alles Sprechen in Sätzen'), in the introduction to his posthumously published volume *Die Syntax des einfachen Satzes im Indogermanischen*. Besides the history of ancient languages, some more recent changes also attracted his attention, most notably the creation of dummy subjects in the modern Germanic languages and in French, to which he devoted an essay (*Der Ursprung des Scheinsubjekts 'es' in den germanischen und romanischen Sprachen*).

Brugmann's work on syntax also shows that, in spite of his central concern with language change, he could also give insightful synchronic descriptions of linguistic phenomena. This is mostly shown in his analysis of word order in the ancient languages, based on what could be termed a functional view of sentence structure. Here, as well as in his assessment of the motivation of language change, Brugmann stressed the importance of the psychological dimension of language use, thus highlighting the role of the speaker in the process of language structuring.

Biography

Karl Brugman was born in Wiesbaden (Germany), on March 16, 1849 as Karl Friedrich Christian Brugman (his last name was changed to Brugmann in 1882). He studied in Halle and Leipzig, and received his Ph.D. Leipzig (1871) for his thesis on suppletion in Greek. From 1872 to 1877, he worked as a teacher in Wiesbaden and Leipzig, habilitated in Leipzig in 1877, and was immediately appointed extraordinary professor.

From 1874 to 1877, he was ordinary professor in Freiburg i. Br. and in 1887 he worked as ordinary professor in Leipzig, where a new chair of 'indogermanische Sprachwissenschaft' (Indo-European Linguistics) had been created specifically for him. He held this position until his death, in Leipzig, on June 29, 1919. He was the founder and editor of 'Indogermanische Forschungen' (founded in 1891) and the first chairman of the Indogermanische Gesellschaft in 1912.

References

Brugmann, Karl. 1917. Der Ursprung des Scheinsubjekts *es* in den germanischen und romanischen Sprachen. Sitzungsberichte der sächsischen Gesellschaft der Wissenschaften, Vol. 5.

———.1876. Nasalis sonans in der indogermanischen Grundsprache. Studien zur griechischen und lateinischen Grammatik 9.

———.1876. Zur Geschichte der stammabstufenden Deklinationen. Studien zur griechischen und lateinischen Grammatik 9.

———.1925. Die Syntax des einfachen Satzes im Indogermanischen. Berlin – Leipzig: De Gruyter.

Brugmann, Karl, and Berthold Delbrück. 1897–1916. Grundriss der vergleichenden Grammatik der indogermanischen Sprachen, 2nd edition. Straßburg: Trübner, reprinted. Berlin: De Gruyter, 1930 and 1967.

Brugmann, Karl, and Hermann Osthoff. 1878 Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen, Vol. 1. Leipzig: Hirzel.

Collinge, N. E. 1985. The laws of Indo-European. Amsterdam, Philadelphia: Benjamins.

Curtius, Georg. 1885. Zur Kritik der neuesten Sprachforschung. Leipzig: Hirzel.

Morpurgo Davies, Anna. 1986. Karl Brugmann and late nineteenth-century linguistics. Studies in the history of Western linguistics. In honour of R.H. Robins, ed. by Th. Bynon and F. R. Palmer. Cambridge: Cambridge University Press.

Morpurgo Davies, Anna. 1996. Nineteenth century linguistics. London: Longman (History of linguistics, Vol. 4, ed. by, Giulio Lepschy).

Sebeok, Thomas A. (ed.) 1966. Portraits of linguists. Bloomington, London: Indiana University Press.

SILVIA LURAGHI

Bühler, Karl

The psychologist and philosopher Karl Bühler started as a member of the Würzburg school of 'Denkpsychologie', where Oswald Külpe was his teacher. Bühler later followed Külpe to Bonn and Munich. In the 1930s, he built up the Psychological Institute of Vienna University, which became one of the most flourishing psychological institutes in Europe. Together with his wife Charlotte, who also worked as a psychologist, he emigrated to Norway and later to the United States when the Nazis had occupied Austria. He spent his last years as a professor of psychiatry in Los Angeles.

Bühler contributed in many ways to the social and cognitive sciences. At the beginning of his career, he worked on the psychology of thinking. Next, he studied human perception and developed a new notion of Gestalt psychology, which he understood as a competitor to the Berlin schools view of Gestalt, as exemplified by Kurt Koffka and Max Wertheimer. Another important part of Bühler's work in collaboration with Charlotte Bühler was developmental psychology. Bühler wrote the most read German textbook on the issue, titled 'Die geistige Entwicklung des Kindes' (1918). Bühler's treatment of the mental development of the child shows a strong concern for the cognitive aspects of language. This research perspective developed into Bühler's 'Sprachtheorie' (1934), which is one of the most important forerunners of semiotics and contemporary cognitive linguistics.

Bühler's most remembered contribution to linguistics is his organon model. See the figure, adapted from 'Sprachtheorie' (1934).

Communication takes place between the speaker (sender) and the listener (receiver), and it is about the objects around us. In Bühler's view, language is an 'organum' or tool for one persons communicating with another about the world. The three main functions of language that Bühler distinguishes in his organon model are *Darstellung* (representation, of states of affairs), *Ausdruck* (expression, of the sender's feelings), and *Appell* (appeal, to the receiver). All functions exist in every single utterance. However, usually one prevails. When the focus is on the feelings of the sender, the expressive function of communication dominates. An object-oriented communication is very neutral or representative. If the focus is on the receiver, we deal with an appeal. The circle in the middle of the illustration above symbolizes the concrete, sensibly given sound. The overlapping triangle symbolizes the meaning of the sound and its *Gestalt*-like features. In
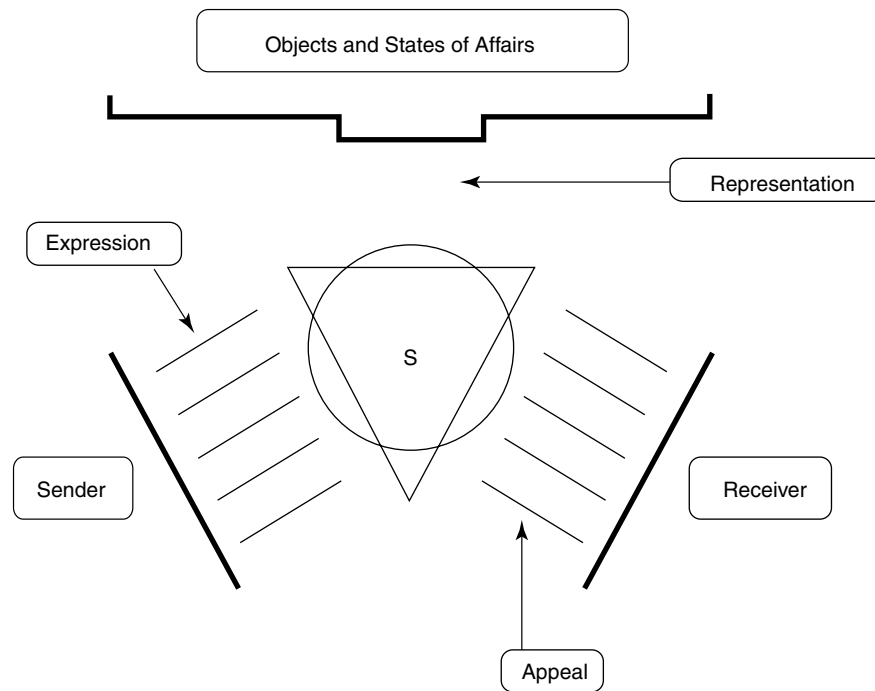
those places where the circle is larger than the triangle, the sound (circle) contains information that lacks meaning (triangle). Where the triangle is larger than the circle, there is a meaning (triangle) lacking an expression in sound (circle). Both phenomena occur all the time in everyday communication. The linguistic sign itself is a symptom, a signal, or a symbol. If it is a symptom, it reveals the interiority or consciousness of the sender. If the sign is a signal, it is directed to the behavior of the receiver. Signs that are mere bearers of information about the states of affairs are symbols, just representing the objects themselves. Roman Jakobson expanded Bühler's model and assumes six functions of language, adding the poetic, phatic, and metalingual function to Bühler's three functions. Karl Popper, who was a student of Bühler's in Vienna, proposes an additional argumentative function.

According to Bühler, the lexicon of a language can be split into the field of symbol words (*Symbolwörter*) and the field of deictic words (*Zeigwörter*). As for deixis, Bühler points out that the listener starts an orientation procedure when the speaker uses a deictic expression.

Bühler shows that the reference is not only based on a sensual perception (when seeing the pointing finger, e.g.). He argues that the pointing finger is the natural instrument of visual demonstration, but it may be replaced by other pointing aids. However, some pointing needs to be done. Bühler's analysis covers the deixis of pronouns ('I', 'you', 'he') and the deixis of time and place expressions ('now', 'here', 'there').

The other field of language, the symbol field, differs radically from deixis. The symbol field is the place for the representational function of language. A symbol bears an invariable content, a content that is independent of the actual situation. The denotation of a deictic (or indexical) expression like 'I' varies from speaker to speaker and from situation to situation. That is the reason why Roman Jakobson calls these expressions 'shifters'. The denotation of a symbolic expression does not vary or shift. Instead, it rigidly denotes always the same thing, like for instance 'Downing Street 10'.

Karl Bühler's *Sprachtheorie* is an important work of twentieth-century linguistics. A historical reevaluation of his work as a whole began in the 1970s. His unpublished works (lectures, manuscripts) are also collected now and edited. It is striking how Bühler's work still stimulates contemporary linguistic research.

Biography

Karl Bühler was born in Meckesheim, Germany on May 27, 1879. He studied human medicine and philosophy at the Universities of Freiburg and Strasbourg. He received a Ph.D. in human medicine (1903, Freiburg) for work related to color perception, a Ph.D. in philosophy (1904, Strasbourg) for work related to Henry Home, and Habilitation (1907, Würzburg) for work on cognitive science. He was assistant professor at the Universities of Würzburg (1906–1909) and Bonn (1909–1913), associate professor at the University of Munich (1913–1918), full professor for philosophy and pedagogy at the Technical University of Dresden (1918–1922), and full professor for psychology at the University of Vienna (1922–1938); he

emigrated to Norway and later to the United States when the Nazis occupied Austria. He was professor at St. Scholastica Duluth College and St. Thomas College at St. Paul, Minnesota (1940–1945), and professor of psychiatry at the University of Southern California, Los Angeles (1945–1955). He died in Los Angeles on October 24, 1963.

References

Brunswick, Egon (ed.) 1929. Beiträge zur Problemgeschichte der Psychologie. Festschrift zu Karl Bühler's 50. Geburtstag. Jena: Fischer.

Bühler, Karl. 1918. Die geistige Entwicklung des Kindes. Jena: Fischer.

———.1927. Die Krise der Psychologie. Jena: Fischer.

———.1933. Ausdruckstheorie. Das System an der Geschichte aufgezeigt. Reprinted with an introduction by Albert Wellek, Stuttgart: Gustav Fischer, Verlag, 1968.

———.1934. Sprachtheorie. Jena: Fischer.

———.1960. Das Gestaltprinzip im Leben der Menschen und der Tiere. Bern: Huber.

Eschbach, Achim (ed.) 1988. Karl Bühler's theory of language. Amsterdam: Benjamins.

Fiske, John. 1987. Introduction to communication studies. London: Methün.

Hülzer, Heike. 1989. Karl Bühler und Wilhelm Stählin: Psychologische Fundamente der Metapherntheorie im ersten Drittel des 20. Jahrhunderts. Münster: Nodus.

Innis, Robert E. 1982. Karl Bühler. Semiotic foundations of language theory. New York: Plenum Press.

MONIKA RATHERT

See also **Jakobson, Roman; Semiotics**

Burmese

Burmese is the national language of the country of Burma (also known as Myanmar), located to the southwest of China and to the east of Bangladesh and India. It is estimated that 68% of the population, or between 22 and 31 million people speak Burmese, ranking it 43rd in the world by population. Burmese is directly related to Tibetan and indirectly related to Chinese. It belongs to the Tibeto-Burman language branch of the larger phylum of Sino-Tibetan. Within the Tibeto-Burman grouping of languages, Burmese belongs to the Loloish (Yi) Branch and shares typological characteristics with lesser-known languages of western China, Thailand, Laos, and Vietnam. Some of the better-known minority languages related to Burmese are Lisu, Akha, and Lahu that are found spread across four Southeast Asian countries.

As a result of internal political changes since 1990, an alternate name of *Myanmar* has been used for both the Burmese language and the country of Burma, but most linguists continue to refer to the language as Burmese. Motivation for the name change is unclear, since the name of the language and country has not changed in Burmese itself. There are two speech levels in Burmese. In the more colloquial, common, spoken form of the language, the name for the country and speech is a form of the word *băma*, from which the English form 'Burma' is derived. In the literary, more formal Burmese, the same name is *myanma.*

There are six to ten recognized regional dialects of Burmese, some of which are not mutually intelligible with Standard Burmese. Standard Burmese is spoken in the highly populated, central Irrawaddy River flood plain and delta regions. Other varieties of Burmese, namely Northern Burmese, Arakanese (Yanbye, Chuangtha), Intha, Danu, Taungyo, Yaw, Tavoyan, and Magui, are found in geographically isolated areas blocked by either mountains or larger bodies of water, such as along the long peninsula toward Malaysia or across the Pegu mountains of Central Burma proper, or the more effectively isolating Arakan mountain range in southwestern Burma.

Ancestors of the modern Burmese-speaking people migrated during the second-to-ninth century CE from the Tibetan plateau southward into areas of southwestern China. Following riverine routes along the eastern end of the Himalayan mountains, they eventually descended into the hot, dry regions of the Irrawaddy River flood plain between 800 and 1000 CE. An earlier migration of a Tibeto-Burman speaking population, the Pyu, preceded the Burmese by 400–600 years. When the Burmese migrated to the plains, the Pyu civilization was in decline. Other language contact with Mon-Khmer-speaking populations, principally the Mon, resulted in significant cultural and linguistic changes for the Burmese. The Mon had already successfully developed an advanced Buddhist civilization, which the Burmese eventually emulated, adopting similarly Indic-influenced systems of government, kingship, religion, art, architecture, and a South Indian variety of the Brāhmī alphabet. The first record of writing Burmese was found on a stone inscription dated 1113 CE. The influence of Mon (and Pali) is found lexically in borrowed words for cultural and technological items, as well as common, everyday idioms. Phonologically, Mon's pervasive

influence is marked in the Burmese tonal system, which has adopted voice phonation or register as additional tonal features in addition to simple pitch.

Significance of Burmese

Burmese is one of the two oldest written languages of the Tibeto-Burman language family. The other is Tibetan. These two languages provide an invaluable window on the language and culture of an ancient linguistic world. Without some written forms of ancient languages, the reconstructed genetic relationships proposed between living languages would be far more speculative. Burmese texts also provide an insight into the political and cultural relationships of a Southeast Asian world shortly after the first millenium. Stone inscriptions are the principal media to have survived. The most famous is the Myazedi inscription dated 1113 CE, which records the same text in four languages: Pyu, Burmese, Mon, and Pali.

The Burmese alphabet is adapted from the Mon script and was one of the war trophies resulting from the conquest of the Mon empire in 1044 CE. The distinctive beauty and harmony of the alphabet is displayed in each character shape, variations of a circle. No word space is used; hence, each word is connected to the previous in a continuous right-to-left flow until a semantic juncture is encountered. A brief space is then used, followed by another long string of elegant circles.

Literary Form and Colloquial Language

The term *diglossia* describes a situation where two extremely different forms of the same language are used for different social functions within a society. This situation arose because the ancient written form exerted a conservative force restricting natural language change. Over the centuries, the written form changed little, while the spoken language more freely followed regular sound changes. The extent to which the written form has constrained changes in the colloquial form is unknown.

Literary Burmese, also called Written Burmese, differs from Colloquial Burmese (Spoken) in style, grammar, lexicon, and grammatical particles, which are similar to English prepositions. The most obvious difference between Literary Burmese and Colloquial Burmese is seen in a unique set of postpositional particles, which serve as spatial, temporal, and logical operators orienting nominal units within the sentence. They also serve as case markers indicating sentence relations such as semantic subject, object, and indirect object. Each style has its own unique set of postpositionals.

In some cases, a historical relationship exists between the particle sets, and in other cases a replace-

ment form is used while maintaining the grammatical function. Lexical forms have been recruited over time to serve standard grammatical functions, demonstrating the flexibility of Burmese semantics and the stability of the grammar.

Dictionaries and Grammars

The first English–Burmese dictionary was compiled by Adoniram Judson, a missionary linguist, during the middle of the nineteenth century and is still used today. In 1993, the Myanmar Language Commission produced a highly popular Myanmar–English Dictionary, which includes a history of the Myanmar alphabet. During the 1970s, the Burmese Language Commission released a five-volume, Burmese– Burmese dictionary set.

Numerous grammars of Burmese have been written. The first in English was Adoniram Judson's grammar of 1866. During the British colonial period of 1885–1948, various grammatical studies produced by local and foreign scholars attempted to grapple with Burmese grammatical constructions using Indo-European, or Latinate models. During the 1960s, various revisionist Burmese linguists proposed alternate grammatical categories using only Colloquial Burmese, rather than the higher status, Literary Burmese. This attempt to legitimate modern Colloquial Burmese met with little official acceptance. The School of Oriental and African Studies, University of London, has made a most significant contribution to the study and analysis of colloquial Burmese for English speakers. Okell's Colloquial Burmese grammar volumes and language lessons have made accessible what was once distant and intangible for the language learner. Unpublished doctoral dissertations since the 1980s have further examined the beauty of the Burmese language.

Phonology

Burmese has four lexical tones, which are characterized by both pitch and voice phonation (plain, breathy, and creaky). The syllable consists of two types: a minor and major syllable. Minor syllables occur only with a major syllable, have only a schwa vowel, neutral tone, pattern as only consonant and vowel (CV), and have a reduced inventory of initial consonants that may occur. The major syllable need not occur with a minor syllable, bears one of the four lexical tones (T), patterns with the potential canonical structure of initial consonant (C_i), optional medial consonants (C_m), vowel (V), and optional final glottal or nasalization (C_f). The major syllable canonical shape is $C_i(C_m)V(C_f)T$.

There are 32 possible initial consonants, two medial consonants, ten vowels, and two types of final consonants. The syllable rhyme (VC_f) is historically important

in determining vowel quality changes such as off-gliding, nasalization, and one of the four tones (Table 1).

Some of the more unusual features of Burmese are the voiceless nasals, *hnim* [n̥eĩ] 'to press down', but *nim* [neĩ] 'settle'. The final nasals, while written, are pronounced only as nasalization on the preceding vowel. Final stops, while written, are pronounced as a glottal stop with an abrupt closure and low tone. This has traditionally been called the stopped or checked tone. The voiceless lateral and the aspirated voiceless sibilant are somewhat universally rare sounds. Surprisingly, Burmese shares with English the moderately unusual voiceless and voiced dental fricative.

Grammar

Burmese is a verb-final language. Typical sentence word order is subject, object, verb (SOV).

Subject	Object	Verb
su	*thã mmàng*	*cà-sany*
s/he	rice	eat-realis
	'S/he eats rice.'	

Sentential nominal elements, such as the subject, object, location, or manner, typically occur with a postpositional marker indicating the sentential role of the nominal. Overt marking of subject and object roles is not obligatory in Burmese as in European languages. Often a *postpositional* is used to disambiguate or contrastively emphasize the role of the nominal.

*mranma nong-ngam-**sany***	*tá-khyin-**ká***	*bríthísh-tó-ʔí*	*lak-ʔawk-**só***	*kyà-rawk-**khái-bù-sany***
Burma-country-**topic**	one-time-**past**	British-plural-**genitive**	hand-under-**toward**	fell-arrived past-previously-realis

'At one time Burma fell under subjugation of the British.'

Noun phrases are normally structured with the modifier preceding the head noun. Relative clauses function as a modifier of the head noun but with a postpositional particle marking the relationship between the preceding clause and the following noun. Possessive phrases also function in the same way.

Possessor	Possessed (Head)
bríthísh-tó-ʔí	*lak-ʔawk*
*British-plural-**genitive***	*hand-under*
'subjugation of the British'	

Burmese is one of the languages in the world that requires the use of semantic classifiers when nouns are enumerated. In enumerated noun phrases modifiers follow the nominal head, with the resulting order: head noun, numeral, classifier. This contrary ordering is a kind of appositional phrase in juxtaposition with the nominal phrase.

TABLE 1 Burmese Sounds

Consonants (C_i)

	Labial	Dental	Alveolar	Alveo-palatal	Palatal	Velar	Glottal
Plosive	pʰ p b	tʰ t d				kʰ k g	ʔ
Fricative		θ ð	sʰ s z	ɕ			h
Affricate					tʃʰ tʃ dʒ		
Nasal	m̥ m	n̥ n			ɲ̥ ɲ	ŋ̥ ŋ	
Glide	w̥						
Lateral				l̥ l			

Vowels (V)

i		u
eɪ	ə	oʊ
ɛ	a	ɔ
aɪ		aʊ

Consonants (C_m)

w	ɹ

Consonants (C_f)

Historically as writtern	-m-n-ɲ -ŋ	-p-t-c-k
Phonetic realization	Nasalization	ʔ

Tone (T)

′ (creaky)	(low, plain)	` (high, breathy)	-p, -t, -k (stopped)

Nominal	Numeral	Classifier
nghak	*tac*	*kawng*
bird	one	body (use of animals)

   'one bird'

In the same way, pluralization of nouns also follows the noun.

Nominal (Head)	Quantifier
ʔalup-sǎmà	*myà*
work-person	many

   'workers'

Burmese treats quantification, plurals, and numeral classifiers, uniquely from other types of specification. Plurals are not as common, nor as grammatically required as in English. Rather, plurals and enumerative phrases make a statement about the topic, and give a sense of inclusion, or definiteness, as articles do in English.

The verb phrase as the last element of the sentence carries the bulk of the action and attitude of the speaker. A series of postpositions refer to the attitude of the speaker to the content of the sentence, the attitude to the hearer, the truth value of the sentence content, and intentionality. Other functions marked by postpositions are verbal aspect (punctiliar, process, completed), mood (declarative, two types of interrogative, imperative), and polarity (positive or negative).

O   Verb Phrase

*bà   lup-**yá-sè-sá-lè***
what   do-**must-yet-realis-question**

   'What else must (you) do?'

Preceding the head verb a limited set of modifiers may occur, which serve as intensifiers, comparatives, manner, or directionals that indicate both actual and metaphorical attunement of the action or state of the main verbal proposition. Following the verb head a more open class of verbal auxiliaries may occur. The juxtaposition of such verbs in Burmese seems to have almost an endless creative potential for creating beauty, humor, and powerful description of visual sense and emotional feeling.

O   Verb Phrase

*ʔaphàm   kham-**swà-rá-hra-sany***
capture   undergo-**go-must-pity-realis**

   '(He) was caught, poor fellow.'

Compounding as a process is extremely productive in Burmese, occurring with words, phrases, and clauses. Compounding occurs with both nominal and verbal constructions. The semantics of the compound construct are not necessarily predictable from the lexical parts but form a semantic blending of various selected components of meaning. Burmese has a noted preference for doublets, such that one compound can be elegantly doubled into a compound of two compounds. Then, these double compounds can be further doubled in a more formal, expressive speech style; veritable doubled pleasure of semantic juxtaposition.

References

Becker, Alton L. 1993. The elusive figures of Burmese grammar: an essay. The role of theory in language description, ed. by William A. Foley. Berlin: Mouton de Gruyter.

Okell, John. 1969. A reference grammar of colloquial Burmese, 2 vols. London: Oxford University Press.

———. 1994. Burmese: an introduction to the spoken language, 2 vols. Dekalb, II: Northern Illinois University, Center for Southeast Asian Studies.

———. 1994. Burmese: an introduction to the script. Dekalb, II: Northern Illinois University, Center for Southeast Asian Studies.

———. 1994. Burmese: an introduction to the literary style. Dekalb, II: Northern Illinois University, Center for Southeast Asian Studies.

Roop, D.H. 1972. An introduction to the Burmese writing system. New Haven: Yale University Press.

PAULETTE HOPPLE

See also **Chinese; Diglossia**

Burushaski

Burushaski (also 'Khajuna, Kanjut') is spoken mostly in the Northern Areas of Pakistan. There are two main dialects of Burushaski: Hunza (HB)/Nager (NB) (approximately 45,000 speakers) and Yasin (YB) (also 'Werchikwar') (approximately 15,000 speakers). The Hunza and Nager variants are quite similar to each other, while the Yasin dialect differs from HB and NB in lexicon, phonology, and morphology. Many HB speakers also speak Shina; YB speakers are bilingual in Khowar. Educated Burushaski speakers also speak and read Urdu, the national and link language of Pakistan. With increasing literacy and bilingualism in

Urdu, many features of Burushaski lexicon and syntax are being influenced by Urdu.

Burushaski is still regarded as a linguistic isolate, although its possible relationship with other languages has long been the subject of intense research. Early attempts to link it with other languages include comparisons with Caucasian and with the Yeneseian languages. Most recently, Čašule (1988) argues that Burushaski is Indo-European, most closely related to the pre-Balkanic languages, especially Phrygian. The most comprehensive grammars of HB are Lorimer (1935a, b, 1938) and Berger (1998); Tiffou (1999) deals with both HB and YB. Anderson (1997, to appear) are English-language analytical summaries of Burushaski phonology and morphology.

Burushaski has recently begun to move toward becoming a literary language. Some writers have chosen to sometimes write in a Roman-based script—in 1980, a Roman-based Burushaski primer was published—while others write in a modified Perso-Arabic script, especially when dealing with religious themes. The Burushaski Research Academy, a local organization, is encouraging the development of writing in Burushaski. Collections of riddles and proverbs have been published in Roman Burushaski, while poetry on religious themes has been published in Perso-Arabic script.

Phonology

Consonants

HB has the consonants shown in Table 1 (Tiffou 1999: 118–9).

In HB, *f* and *x* appear in loan words; *y̱* is a retroflex glide, unique to HB and NB. Except for *y̱*, YB has the same consonants as HB; it also has *x*, corresponding to HB *qh*.

Vowels

Both HB/NB and YB have five basic short vowels, i, e, a, o, u, and two sets of corresponding long vowels (consisting of two time units, or moras), one stressed on the first mora, represented as *áa,* and another stressed on the second mora, represented as *aá*. First-mora stress results in a high-falling pitch, while second-mora stress results in a low-rising pitch contour. Vowel length and moraic stress are contrastive, e.g. *bat* 'stone', *baát* 'cooked rice', and *báat* 'after' (Berger 1998). This places Burushaski in the areal grouping of north Pakistani languages, including at least Shina, Khowar, and Burushaski, with pitch accent systems.

Morphology

Nouns

Burushaski nouns fall into a four-valued gender class system. The four classes are human female (hf), human male (hm), x, and y. The x class includes nonhuman animates and most count nouns, while the y class includes inanimates and mass nouns. The class membership of a noun determines verb and sometimes adjective agreement; in some cases, the stem form of a verb also varies according to the class of the subject or the direct object, e.g. *yáaltas* 'to wash (h or x object) vs. *báaltas* 'to wash (y object), and *wálas* 'to fall (x subject) vs. *balás* 'to fall (y subject). There are numerous plural suffixes in Bur., which are correlated with the class of the noun. For example, *-tiŋ* (hm, hf) in *daraɣá-tiŋ* 'canal overseers'; -ŋ (y) in *hála-ŋ* 'goals (polo)' *-muts* (x) in *baṣá-muts* 'turbans'. Bur. has a suffixal singular article *-an* 'a, one' (from *han* 'one'), and a plural article *–ik;* thus, *hayúr-an* 'a horse'; and *harált-iŋ-ik* '(some) rain-showers' (Berger 1998, I:39, 43).

Nouns have an elaborate case-marking system, consisting of several levels. The basic case endings are

TABLE 1

Manner of articulation	Point of articulation						
	Bilabial	Dental	Palatal	Retroflex	Velar	Postvelar	Glottal
Stops							
Voiceless	p	t		ṭ	k	q	
Voiceless, aspirated	ph	th		ṭh	kh	qh	
Voiced	b	d		ḍ	g		
Affricates							
Voiceless	(pf)	ts	č	c̣			
Voiceless aspirated		tsh	čh	c̣h			
Voiced			j	j̣			
Fricatives							
Voiceless	(f)	s	š	ṣ			h
Voiced		z	(ž)		γ		
Nasals	m	n					
Continuants	w	l	y	y̱			
Vibrant		r					

TABLE 2

Case	2 'man' (hm)	3 'woman' (hf)	4 'horse' (x)	5 'sword' (y)
Nominative/ absolute	*hir*	*gus*	*haɣur*	*ɣaténç*
Oblique	*hir*	*gus-mu*	*haɣur*	*ɣaténç*
Ergative	*hir-e*	*gus-e*	*haɣur-e*	*ɣaténç -e*
Genitive	*hir-e*	*gus-mu*	*haɣur-e*	*ɣaténç -e*
Dative	*hir-ar*	*gus-mo-r*	*haɣur-ar*	*ɣaténç -ar*
Ablative	*hir-tsum*	*gus-mu-* *tsum*	*haɣur* *-tsum*	*ɣaténç* *-tsum*

TABLE 3

	Type I, unstressed	Type I, stressed	Type II	Type III
1	*a-*	*á-*	*á-*	*áa-*
2	*gu-*	*gú-/-kú-*	*gó-/-kó-*	*góo-/* *kóo-*
hm	*i-*	*í-*	*é-*	*ée-*
hf	*mu-*	*mú-*	*mó-*	*móo-*
x	*i-*	*í-*	*é-*	*ée-*
y (sg/pl)	*i-*	*í-*	*é-*	*ée-*
1 (pl)	*mi-*	*mí-*	*mé-*	*mée-*
2 (pl)	*ma-*	*má-*	*má-*	*máa-*
hm, hf, x (pl)	*u-*	*ú-*	*ó-*	*óo-*

illustrated in Table 2 for one noun of each class (Anderson, to appear).

The ergative case is used for the subjects of transitive verbs, such as 'hit', while the nominative/absolutive case is used for the subject of intransitive verbs, such as 'go'. For some speakers, though, the nominative/absolutive can be used in the future and the present of transitive verbs. The genitive indicates possessors; the dative marks indirect objects, some direct objects, and direction toward something. The ablative indicates direction away from something and appears in comparative constructions. Two examples follow.

> *jáa-r*   *madát*   *a -čh -í*
> I(OBLIQUE)-DATIVE   help   I-give
> 'help me'
> (Berger 1998,I:69)

> *uŋ -tsum*   *je kam*   *a- p -a*
> you-ABLATIVE   I   small   not-be-I
> 'I am not smaller/less than you.'   (Anderson, to appear)

In addition, there are several locative cases, and a fairly large set of 'relational nouns'. These words are nominal in origin, but are undergoing the process of grammaticization into postpositions.

Pronouns

Pronominal prefixes are a central element of Burushaski morphology. In general, these prefixes are used to indicate an affected, animate entity. They are obligatory with certain nouns, e.g. names for body parts and kinship terms, indicating inalienable possession. For example, *-me* 'tooth' cannot stand alone; it must include a prefix indicating 'whose tooth', e.g. *gu-me* 'your tooth'. Some adjectives and postpositions, and a few numbers require pronominal prefixes, as do many verbs. For example, *-apat* 'side of body' is a body part term requiring a pronominal prefix, from which the adjective *-pačim* 'being alongside of __', *go-pačim* 'by your side' (Bashir 1985:3). With verbs, pronominal prefixes index the subject of some intransitive verbs, the object of some transitive verbs, or the indirect object or causee of some ditransitive or

causative verbs. There can be more than one pronominal prefix in a sentence. For example, in the following sentence, the first-person singular pronominal prefix *áa* indexes the affected object 'me', and the second-person singular prefix *kóo* indexes the subject 'you'.

> *áa-lji*   *du-kóo -ṣqalč*   *-um*   *-a*
> me-up.to   *d* -you -overtake(DURATIVE)   -PPL-you
> 'You will overtake me.'
> (Bashir 1985:15, from Lorimer 1938:119)

There are four series of these prefixes (Berger 1998:91). See Table 3.

Independent pronouns also exist. They are:

>   Sg.   Pl.
> 1. *je, já* 'I'   *mi* 'we'
> 2. *un* 'you'   *ma* 'you'
> 3. *iné* (h) 'he, she'   *ué* (h) 'they'
>   *isé* (x) 'it'   *itsé* (x) 'they'
>   *ité* (y) 'it'   *iké* (y) 'they'

Adjectives

Some plural adjectives agree in class with nouns, e.g. *burúm* 'white (sg.)', *burúm-išo* 'white-x pl', *burúm-iŋ* 'white-y pl' (Berger 1998:47). Comparative and superlative constructions are formed analytically with the ablative case.

Verbs

Burushaski has a relatively small number of simple verbs; the majority of verbs consist of a noun or adjective combined with one of a small set of basic verbs (sometimes called "light verbs"), which function to change the noun or adjective into a verb. Of these, the most frequently used are *mánas* 'to be', which forms intransitive verbs, and *étas* 'to do, make', which forms transitives. For example, *naṭ étas* 'to dance', *kárpaṭ étas* 'to quarrel', *kárpaṭ mánas* 'to quarrel'.

The Burushaski verb is agglutinative; that is, it is formed of several separate meaningful units (morphemes), each of which corresponds to a specific meaning element. The general structure of a verb form can be schematized as follows (Anderson, to appear).

These positions are ordered from left to right in the verb form (See Table 4).

One example of a fairly complex verb form, in itself a complete sentence, is given here. Position numbers and morpheme names are indicated; note that not all the positions are filled in any one verb.

TABLE 4

Position	Morpheme	Function
–4	**a-, oó-**	Forms a negative verb
–3	**d-, n-**	Prefixes/preverbs. *d-* indicates (abstract) motion in the direction of speaker or terminus (Bashir 1985) but is no longer productive. *n-* marks the conjunctive participial (also 'converb', 'absolutive') form of the verb.
–2	**Pronominal prefixes**	Index the subject, indirect object, or direct object of a sentence, depending on the transitivity of the verb.
–1	**Causative morpheme -s-**	Indicates either a causative verb, or a benefactive meaning.
0	**verb root**	Contains the basic meaning of the verb.
+1	**plurality marker -ya-**	Marks an action as distributed in some way, either by having a plural subject or by being performed on multiple occasions or in multiple places.
+2	**durative, nonpast participle marker: c, š, c̣, j, y;**	Markers of durative (cf. 'present') base. They indicate that an action is durative, i.e. has temporal extent.
+3	**1st person singular subject -(y)a-**	Marks only first-person singular subjects; thus, position 3 and position 5 cannot both be filled.
+4	**(a) m- participle marker; (b) infinitive -as; (c) conjunctive participle suffix -n; (d) optative -?; (e) optative -áa.**	These all mark nonfinite verb forms, or optatives.
+5	**(a) subject suffixes other than 1st sg.; (b) imperative endings; (c) auxiliary verb ba-.**	The auxiliary *ba-* 'be' forms complex verb tenses and itself has personal endings.
+6	**(a) interrogative; (b) case ending**	(a) Marks the sentence as a question. (b) Enables a sentence to be embedded in a more complex sentence.

a	-tú	-ku	-man	-um	-a
NEG	-d	-you(sg).	-be.born	-NON-DURATIVE-	you(sg)
[–4	–3	–2	0	+4	+5]

'You weren't born'
(Berger 1998,I:91)

Tense forms fall into two categories: those formed on the nondurative (sometimes called 'past') stem and those formed on the durative (sometimes called 'present') stem.

Sentence Structure

Basic Word Order

Burushaski has postpositions, and adjectives precede nouns. The basic word order of a simple sentence is SUBJECT – OBJECT – VERB. For example:

híre   gus   mu-yeétsimi
man   woman   saw
'The man saw the woman.'
(Berger 1998,I:177)

Observe that the direct object 'woman' is expressed on the verb by the prefix *mu-*, for human females.

Complex Sentences

In complex sentences, the subordinate clause usually precedes the main clause. The indigenous strategy for forming subordinate clauses involves using one of the nonfinite verb forms. For example, one type of relative clause is formed with the infinitive:

*[buṭ   bulá **dél-as]**   iné   hír-e*
much polo hit-**INFINITIVE** that man
'the man who plays much polo…'
(Berger 1998,I:186)

Another construction uses the -*m* participle (position +4).

*ja baríŋ   **etum**   ne   hir   šua   bái*
I speech **done** that man good   is
'The man with whom I spoke is good.'
(Tiffou 1999:2000)

In addition, some relative clauses are formed with finite verbs and a relative (-like) element. For example, the following YB sentence:

*ámet caγá-ule gambúrimu bién **(te)**   ṣiéli   duá*
that place-in flowers   are (that) beautiful is
'The place where there are flowers is beautiful.'
(Tiffou 1999:201)

The conjunctive participial form, marked by *n/d-*VERB-*n* (Positions: –3, 0, 2) forms many adverbial clauses. For example:

íne hír-e	***d-ítal-in***	*ṣapík*	*ṣími*
that man	*d*-get up(CONJUNCTIVE PARTICIPLE)	bread	ate

'That man got up and ate bread.'
(Berger 1998, I:189)

In many types of adverbial clauses, a case marker follows a nonfinite verb form. For example:

*sén-as-**ar***
say-INFINITIVE-**DATIVE**
'When he said'
(Berger 1998,I:190)

Subordinating constructions augmented by the complementizer *ke* 'that' (probably an influence from Khowar and/or Urdu) are increasingly frequent. For example,

je girámar	***ke***	*in*	*ními*
I village-to	when	came	he-went

'When I came to the village, he left.'
(Berger 1998, I:191)

References

Anderson, Gregory D.S. 1997. Burushaski phonology. Phonologies of Asia and Africa (including the Caucasus),Vol. 2, ed. by Alan S. Kaye, 1021–41. Winona Lake, IN: Eisenbrauns.

_____. to appear. Burushaski morphology. Morphologies of Asia and Africa. Winona Lake, IN: Eisenbrauns.

Bashir, Elena. 1985. Toward a semantics of the Burushaski verb. Proceedings of the conference on participant roles: South Asia and adjacent areas, ed. by Arlene R.K. Zide, David Magier, and Eric Schiller, 1–32. Bloomington, IN: Indiana University Linguistics Club.

_____. 2000. A thematic survey of Burushaski research. History of Language 6(1). 1–14.

Berger, Hermann. 1974. Das Yasin-Burushaski (Werchikwar): Grammatik, Texte, Wörterbuch. Neuindische Studien,. Bd. 3. Wiesbaden: O. Harrassowitz.

Berger, Hermann, with collaboration of Nasiruddin Nasir Hunzai. 1998. Die Burushaski-Sprache von Hunza und Nager. Neuindische Studien, Bd. 13. Wiesbaden: Harrasowitz.

Čašule, Ilija. 1998. Basic Burushaski etymologies: the Indo-European and Paleo-Balkanic affinities of Burushaski. LINCOM Etymological Studies, Vol. 01. München: LINCOM Europa.

Lorimer, D.L.R. 1935–1938. The Burushaski language. Vol. I, Introduction and Grammar. Vol. II, Texts and translations. Vol. III. Vocabularies and index. Publikationer Serie B: Skrifter - Instituttet for sammenlignende kulturforskning. 29, 1–3. Oslo: Instituttet for Sammenlignende Kulturforskning, H. Aschehoug.

———. 1962. Werchikwar English vocabulary; with a few Werchikwar texts, Vol. 51. Instituttet for sammenlignende kulturforskning. Serie B: Skrifter. Oslo: Norwegian Universities Press.

Tiffou, Étienne. 1999. Parlons Bourouchaski. Paris: L'harmattan.

Tiffou, Étienne, in collaboration with, Yves-Charles Morin, Hermann Berger, D.L.R. Lorimer, et al. 1993. Hunza proverbs. Calgary: University of Calgary Press.

Tikkanen, Bertil. 1995. Burushaski converbs in their South and Central Asian areal context. Converbs in cross-linguistic perspective: structure and meaning of adverbial verb forms-adverbial participles, gerunds. Vol. 13, Empirical approaches to language typology, ed. by Martin Haspelmath and Ekkehard König, 487–528. Berlin: Mouton de Gruyter.

Willson, Stephen R. 1999. Basic Burushaski vocabulary. Studies in Languages of Northern Pakistan, Vol. 6. Islamabad: National Institute of Pakistan Studies, Quaid-i-Azam University and Summer Institute of Linguistics.

ELENA BASHIR

C

Canada

Canada, whose name derives from the Iroquois word [ganá:da] 'settlement', is a confederation of ten provinces and three territories in northern North America. Formerly a British colony, Canada was created by the British North America Act (1867), which joined the provinces of Upper Canada (Ontario), Lower Canada (Québec), New Brunswick, and Nova Scotia into a single dominion. Other provinces joined the confederation at later dates: Manitoba (1870), British Columbia (1871), Prince Edward Island (1873), Alberta (1905), Saskatchewan (1905), and Newfoundland and Labrador (1949). In 1875, the Yukon and Northwest Territories were incorporated into the confederation, and the territory of Nunavut was created in 1999. Although Canada is the second-largest country in the world, encompassing an area of 9,922,385 km^2, it has a population of only 30,007,094 (according to the 2001 census), most of which is concentrated along its southern border with the United States.

In light of its vast size, Canada's present multilingual situation should not be surprising. A distinction is normally made between the 'founding languages' (English and French), the aboriginal (or First Nations) languages, and the 'heritage' (or 'international') languages of immigrant groups, which are neither Anglophone (English speaking) nor Francophone (French speaking). As the following figures from the 2001 census demonstrate, English is claimed by the majority of Canadians both as their mother tongue and as the language they use at home on a daily basis, with French as the second-largest language.

	Home language		Mother tongue	
English	20,011,535	67.5%	17,521,880	59.1%
French	6,531,375	22.0%	6,782,320	22.9%
Nonofficial language	3,096,110	10.5%	5,334,770	18.0%

The British North America Act provided for the use of both English and French in Parliament and the Québec legislature, provisions that were later extended to other provinces and territories, but in practice most of the provincial legislatures and much of the federal government were English dominant until the 1970s. Since the adoption of the Official Languages Act (1969), Canada has maintained a policy of official bilingualism in English and French at the federal level, although all provinces except New Brunswick are still officially monolingual in either French (Québec) or English. Unlike the 'melting pot' policy of the United States, the Canadian policy since 1971 has been one of 'bilingualism within a multicultural framework'. This framework attempts to promote conditions for stable bilingualism, develop the prestige of French in Canada, and protect individual and minority language rights. These rights were further enshrined in the Constitution and the Charter of Rights and Freedoms (1982). The effect of federal policy has been to make competence in both official languages a valuable asset in politics, law, culture, and business. French immersion courses have increased official-language bilingualism among Anglophones, especially in Québec, although Francophones are still the most bilingual nationally.

Canada's first inhabitants arrived from Siberia across the Bering land bridge between 7,000 and 12,000 years ago, bringing languages belonging to linguistic families that are now spread throughout the hemisphere. Although there is disagreement about the exact historical affiliation of these families, Canadian aboriginal languages are usually grouped into 11 families and isolates, most of which are also spoken in the United States, Greenland, or Siberia. Varieties of Inuktitut cover a wide area across the north, from the Yukon to Labrador. British Columbia is home to the widest array of languages, including Kutenai and Tlingit, the Haida dialects, and the Tsimshian, Wakashan, and Salish families. The Athabaskan family stretches across the west from British Columbia to Manitoba and north into the Yukon and Northwest Territories. The Algonquian family, which includes Blackfoot, Cree, Ojibwa, Micmac, and possibly Beothuk (an extinct language of Newfoundland), is the most widespread, with languages spoken from Alberta to Labrador, as well as in New Brunswick and Nova Scotia. Michif, a mixed language that arose from intermarriage among speakers of Cree and French, is spoken in Manitoba. The Dakotan languages (a branch of the Siouan family) are spoken in Manitoba, Saskatchewan, and Alberta, and a number of Iroquoian languages are spoken in Ontario and Québec. According to the 2001 census, the aboriginal languages with the largest number of speakers are Cree, Inuktitut, and Ojibwa.

When the Europeans arrived in North America, there were up to 300 languages spoken in what is now Canada, although nowadays there are fewer than 100. This attrition, the result of infectious diseases brought by Europeans, genocide, and government policies, has only recently begun to be addressed. After Confederation, little attention was given to the linguistic rights of aboriginal peoples, most of whom did not have full citizenship. Until the 1960s, the federal government pursued an assimilationist policy, confining aboriginal peoples to reserves and giving educational authority to religious groups, who enforced instruction in the majority language (English or French). More recently, aboriginal groups such as the Assembly of First Nations have called for a greater degree of self-determination, closing down the residential schools and transferring educational responsibility to local authorities. Some aboriginal languages now have official status in the Northwest Territories and Nunavut. Although federal and provincial funding for aboriginal-language instruction has helped slow language loss to some degree, these programs are inconsistent across the country and do not always promote fluency.

The varieties of French spoken in Canada can be traced to two colonies founded in the early seventeenth century: Acadia (1604; now Nova Scotia, New Brunswick, and Prince Edward Island) and New France (1609, now Québec). Although Acadia was ceded to the British in 1716 and many of the Acadians were later deported, Acadian French survived and spread to Newfoundland, the Îles de la Madeleine, and the Gaspé Peninsula (Québec). By 1762, the British had also conquered New France, but in this case they guaranteed the colonists' right to the French language. Québec French is nowadays the mother tongue of most inhabitants of the province and, because of westward migration in the nineteenth and twentieth centuries, served as the basis of the French spoken in Ontario, Manitoba, and the western provinces. Because of their different patterns of settlement and subsequent development, Acadian and Québec French show a number of distinctions from each other in grammar and pronunciation, especially in their vowel systems and verbal inflection. Both differ from European French, most noticeably in vocabulary and pronunciation, especially Acadian French and colloquial Québec French, which is also known as *joual* (from *cheval* 'horse'). Although there is a perception of massive and pervasive influence from English, anglicisms and English borrowings in fact account for a very small percentage of the total vocabulary.

Although French was originally the majority language of Canada, concentrated in Lower Canada and Acadia, the influx of Anglophone settlers to Upper Canada and the Atlantic provinces resulted in a national decline in the Francophone percentage: from 50% to approximately 25% in the nineteenth century. In addition, in all provinces except Québec and New Brunswick, the Francophone population has gradually been shifting to English. These patterns have decreased the presence of French in Canada while simultaneously increasing its territorialization. By the middle of the twentieth century, these factors, plus the declining birthrate among Québec Francophones, led to the perception of a 'doomsday scenario' in which French would eventually disappear from Canada. In 1976, this perception prompted the newly elected nationalist Parti Québécois government of Québec to pass Bill 101, the *Charte de la Langue Française*. Bill 101 ensures the rights of the Francophone majority by making French the sole official language and requiring all non-Anglophone immigrants to attend French-language schools, all public signs to be in French, and all businesses to conduct their affairs in French. This policy is implemented by three boards responsible for codifying and developing French, monitoring its status, and enforcing the language laws. Despite studies

indicating that Québec Francophones still do not feel that the French language in the province is secure, the language laws have increased the status of French in Québec to the point that it is possible to live and work entirely in French, even in the bilingual city of Montréal, and the percentage of Québécois reporting French as their mother tongue has remained stable (81%) from 1986 to 2001. However, the prestige of English as the language of North American business and culture poses a continuing threat to the maintenance of French in Canada.

The English presence in Canada dates from 1497, when Newfoundland was claimed for the British Crown. Newfoundland was settled between the sixteenth and eighteenth centuries almost entirely from Ireland and southwestern England and remained relatively isolated until World War II, making Newfoundland English distinctive in pronunciation, grammar, and vocabulary from varieties of Canadian English spoken elsewhere. Present-day Standard Canadian English is usually traced to the two waves of immigration of United Empire Loyalists (1776–1793), Americans loyal to the British monarch who fled the American Revolution: one wave came from coastal New England (Connecticut, Massachusetts, Maine, and Rhode Island) and settled in Nova Scotia and New Brunswick, whereas another came from the midland (Pennsylvania, New Jersey, New York, and Vermont) and settled in Ontario. The Loyalists thus served as linguistic models for subsequent settlers recruited by the British government in the nineteenth century from England, Scotland, and Ireland to augment the Anglophone population and supply agricultural and construction workers. Because the western provinces were settled largely by migration from the east, Standard Canadian English is nowadays considered extremely uniform across the country, without the marked regional standards characteristic of the United States. Despite this apparent uniformity, several nonstandard regional varieties persist, such as Newfoundland English, the African-Canadian English spoken by the descendants of Black Loyalists in Nova Scotia, Ottawa Valley English, and the noticeably distinctive pronunciation of certain vowels, especially before [r], in the Atlantic provinces. Although the shared origins of Canadian English and U.S. English make them grammatically very similar, there are differences in pronunciation, vocabulary, and spelling practices. Most noticeable in pronunciation is the raising of the first part of the vowel in words such as 'house', which Americans often hear as 'hoose'. There is considerable variation among Canadians between British and American spelling standards in words such as colour/color and cheque/check.

After 1860, immigration was widened to European countries outside of the British Isles, drawing settlers from Germany, Italy, the Scandinavian nations, and Ukraine. Such settlers, who increased the population of the eastern cities and were instrumental in settling the western provinces, tended to give up their ethnic languages in favor of English, further increasing the Anglophone majority. After 1910, immigration widened to settlers from Russia, Iceland, Finland, Poland, Greece, and the Netherlands. Since World War II, immigrants have also arrived from southeastern Europe, eastern Africa, India, Pakistan, Korea, Vietnam, China, Hong Kong, and the Philippines. With the adoption of multiculturalism as a federal policy in 1971, immigration was seen as a response to falling birthrates and a higher average age in the population. The annual rate of immigration has varied considerably, but currently Canada accepts approximately 200,000 immigrants every year.

The short-term effect of immigration has been to increase the number of nonofficial languages spoken: in the 2001 census, almost 100 nonofficial languages were claimed as a mother tongue by 18% of the population. The ten largest groups are (in descending order) Chinese, Italian, German, Punjabi, Spanish, Portuguese, Polish, Ukrainian, Arabic, and Tagalog (Pilipino). Immigrants tend to settle in southern Ontario (especially Toronto), Vancouver, and Montréal, making Canada's urban population increasingly multiethnic. However, because most immigrants can speak at least one official language within a short time of arrival, the long-term effect of immigration is more likely to be social than linguistic. Indeed, although immigration has contributed to social conflict, the ongoing educational and political accommodation of so many different languages within one state reflects favorably on the Canadian linguistic experiment.

References

Chambers, J.K. 1991. Canada. English around the world: sociolinguistic perspectives, ed. by Jenny Cheshire. Cambridge: Cambridge University Press.
Churchill, Stacey. 1998. Official languages in Canada: changing the language landscape. Ottawa: Canadian Heritage.
Clarke, Sandra (ed.) 1993. Focus on Canada. Amsterdam and Philadelphia: John Benjamins.
Edwards, John (ed.) 1998. Language in Canada. Cambridge: Cambridge University Press.
Marmen, Louise, and Jean-Pierre Corbeil. 1999. Languages in Canada: 1996 Census/Langues au Canada: recensement de 1996. Ottawa: Canadian Heritage.

JAMES A. WALKER

See also **English; French Language**

Cape Verdean Creole

The Cape Verde islands belong to the northwestern union of archipelagoes in the Atlantic ocean (including the Canary Islands, the Azores, and Madeira) and are situated approximately 450 kilometers from Senegal and Mauritania. The archipelago is composed of two main clusters of islands: the windward islands and the leeward islands, known locally as Barlavento and Sotavento, respectively. The Barlavento islands include Boavista, Sal, São Nicolau, Santa Luzia, São Vicente, and Santo Antão. The Sotavento islands consist of Brava, Fogo, Santiago, and Maio. Some 450,000 Cape Verdeans currently reside in the islands and one million reside abroad (primarily Portugal, the Netherlands, New England, and California).

Although the Portuguese and the Italians officially discovered the islands in approximately 1460, there is cartographic evidence that the Greeks and the Arabs were already aware of their presence as early as in 1413 (Andrade 1996). Furthermore, the Jalofos tribe is reported to have lived on the island of Santiago prior to the arrival of the Portuguese (Brásio 1962).

Strategically located at the crossroads of Europe, Africa, and America, the archipelago played a critical role in the slave trade from the fifteenth to nineteenth centuries. The Portuguese initially brought the slaves to Cape Verde from Guinea, including Senegal and Sierra Leone. However, competition from other western powers reduced the West African source of slaves for Cape Verde to the limits of present-day Guinea-Bissau by the middle of the sixteenth century (Carreira 1982). Historical sources state that most of the slaves in Cape Verde came from the tribes of the Mandingos, Balantes, Bijagos, Feloupes, Beafadas, Pepels, Quissis, Brames, Banhuns, Peuls, Jalofos, Bambaras, Bololas, and Manjakus (Brasio 1962). The white settlers came from Algarve and Alentejo in Portugal, and also included Jews, Spaniards, Italians, and French (Martinus 1996).

A number of morphophonological and syntactic features distinguish the Barlavento varieties from their Sotavento counterparts, which can, to a certain extent, be related to the greater influence of African languages in the Sotavento varieties. This disparity is in part due to the fact that the Barlavento islands were settled a hundred years later than the Sotavento. As a result, the sociolinguistic situation in Cape Verde is fairly complex; it also involves an ideological dimension, as Portuguese (which provided the language with most of its lexicon) is the only official language in the archipelago to date.

In the phonological domain, all dialects of CVC contain the vowels {u, ɛ, a, e, ó, o, i}. The Barlavento group also preserves the Portuguese diphthongs *ei* and *oi*, which have merged with *e* and *o* in Sotavento. (Compare Port. *Manteiga* 'butter' : Sot. *mantega*, Port. *açoitar* 'to whip' : Sot. *sota*.) Sotavento varieties also typically denasalize word-final vowels; compare Port. *tambén* 'too' : Sot. *tanbe*. The Sotavento varieties change *v* into *b*, as in Port. *vestir* 'to dress' : CVC *bisti*. The Portuguese palatal liquid *lh* typically becomes the affricate *dj* in CVC: Port. *velho* 'old' : CVC *bedju.*

In the morphological domain, Sotavento differs from Barlavento in making productive use of reduplication, a morphological process also found in African languages, whereby a reduplicated adjective or adverb expresses emphasis, as in *moku moku* 'very drunk' or *faxi faxi* 'very quickly'. Noun reduplication may yield a distributive interpretation, as in *dia dia* 'every day', or may simply lead to a change in meaning, as in *boka* 'mouth' → *boka boka* 'in secret'. Lexical categories may shift grammatical category when reduplicated, as in *mansu* 'quiet' (adjective) → *mansu mansu* 'secrecy' (noun). Reduplicative processes of this type are relatively scarce in Barlavento.

The lexicon of Cape Verdean Creole comes primarily from Portuguese, with a smattering of African forms, such as *banbu* 'to carry a child on one's back' from Manding and Manjaku *banbu* (same meaning), *fepu* 'entirely' from Wolof *fepp* 'everywhere', *lokoti* 'to get out of a hole' from Wolof *loqoti* (same meaning) (Santos 1979). Many of the words of African origin refer to flora and fauna.

The verbal system is fairly intricate. The bare verb stem assumes various functions, depending on whether the verb is stative or nonstative. The bare stem of the verb *ten* 'have, own, possess' has a present tense reading, as in *N ten dos fidju* 'I have two sons'. In contrast, the bare stem of a nonstative verb yields a past tense interpretation, as in *N papia ku bu pai* 'I spoke with your father'. Creole languages typically have preverbal unbound markers expressing Tense, Mood, or Aspect (TMA). Cape Verdean Creole has two TMA markers, *ta* and *sta*, which occur preverbally, as illustrated in *algen ku si boka ka ta era kaminhu* 'one who asks does not get lost' and *pamodi bu sta txora?* 'why are you crying?' The marker *dja* in contrast precedes subject clitics but must follow full noun phrases; compare *Dja'N atxa bonberu* 'I found the exterminator' vs. *N atxa bonberu dja poi nunbru na porta* 'I found that the exterminator had

already put the number on the door'. In contrast to other creoles, Cape Verdean Creole possesses the inflected anterior marker –*ba*, which expresses simple past with stative verbs and pluperfect with nonstative verbs, as in *Mininu ten*ba *febri* 'The child had fever' and *João kume-*ba *tudu galinha* 'John had eaten the whole chicken'.

While the importance of inflectional morphology in creoles has been minimized by a number of creolists, Cape Verdean Creole and a few others display inflectional morphology not only on verb stems, as seen with the –*ba* verbal inflection above, but also on noun stems. CVC shares this feature with a number of other creoles and pidgins, such as Guinea-Bissau Creole, Casamance Creole, Ghanaian Pidgin English, and Nigerian Pidgin English. Plural suffixation in Cape Verdean is a productive process subject to definiteness and the animacy hierarchy. (The animacy hierarchy entails that animate entities are more likely than inanimate objects to carry agreement markers such as the plural suffix.) From a cognitive perspective, animate and/or definite entities are more salient in the speaker's consciousness and hence are more likely to be pluralized (Lyons 1999). The following examples illustrate this feature: *Bu tra-badja ku nha povus* 'you worked with my people', *Ami nha mininus, es fika tudu la pa fora* 'As for me, my children, they are all over the place'. Although the study of large corpora reveals that plural suffixation in monolingual speech is sensitive to variables such as [+/-animate], [+/-human], and [+/-definite], it is important to keep in mind that monolingual speech is not a monolithic whole and that speakers may have multiple grammars. Variation in general has been a central theme in the study of Cape Verdean Creole (Baptista 2002). In addition, social class and education, higher degrees of which tend to be bound up with diglossia and bilingualism (cf. Duarte 1994), combine with other social factors to make the Cape Verdean language quite heterogeneous, as is expected of any natural language.

References

Andrade Silva, Elisa. 1996. Les Iles du Cap-Vert: De la Découverte à l'Indépendance Nationale (1460–1975). Paris: L'Harmattan.

Baptista, Marlyse. 1997. The morpho-syntax of nominal and verbal categories in Cape Verdean Creole. Ph.D. dissertation. Harvard University, Ann Arbor, MI: UMI Dissertation Services.

Baptista, Marlyse. 2002. The syntax of cape verdean creole: the sotavento varieties. Philadelphia, Amsterdam: John Benjamins.

Brásio, António. 1962. Descobrimento, Povoamento, Evangelização do Arquipélago de Cabo Verde. Lisbon: Centro de Estudos Históricos Ultramarinos.

Brito, Ana Paulo. 1967. Dialectos Crioulos-Portugueses: Apontamentos para a Gramática do Crioulo que se Fala na Ilha de Santiago de Cabo Verde. Estudos Linguísticos Crioulos, revised by Adolpho Coelho, ed. by Jorge Morais-Barbosa [originally published in 1887]. Lisbon: Academia International da Cultura Portuguesa.

Carreira, Antonio. 1982. O Crioulo de Cabo Verde: Surto e Expansão. Lisbon: Carreira edition.

Duarte, Dulce. 1994. Crioulo Caboverdiano e Diglossia. Papia 3. 32–45.

Ferraz, Luiz, and Marius Valkhoff. 1975. A comparative study of São-Tomense and Cabo-Verdiano Creole. Miscelanea Luso-Africana: Colectanea de Estudos Coligidos, ed. by Marius Valkhoff. Lisbon: Junta de Investigações Científicas do Ultramar.

Lyons, Christopher. 1999. Definiteness, Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Martinus, Frank. 1996. The kiss of a slave: Papiamentu's West-African connections. Ph.D. Dissertation. University of Amsterdam.

Petit Futé. 1999. Coutry guide: Cap-Vert. Paris: Nouvelles Editions de l'Université.

Santos, Rosine. 1979. Le Créole des Iles du Cap-Vert: Comparaison avec les Langues Africaines. Réalités Africaines et Langue Française 11. Dakar: Centre de Linguistique Appliquée de Dakar.

Veiga, Manuel. 1998. Le Créole du Cap-Vert: Etude Grammaticale Descriptive et Contrastive. Thèse Doctorale. Aix-en-Provence: Université Aix-Marseille.

BERT VAUX AND MARLYSE BAPTISTA

Carib and Cariban Languages

Languages belonging to the Cariban language family are spoken in northern South America, the bulk in northern Brazil, Venezuela, Guyana, Surinam, and French Guiana, with outliers to the west in Colombia and to the south in Central Brazil. The historical literature on the Cariban family names over 100 languages; however, linguistic information (e.g. word lists, brief collections of utterances) exists for maybe half. There is no question that many of the languages spoken at the time of first contact with Europeans have become extinct, but we will never know how many. Currently, some 25 Cariban languages remain, with a cumulative total of between 60,000 and 100,000 speakers; well over half of these speak Carib proper or

one of three closely related languages belonging to a single subgroup (the Pemóng Proper Subgroup), with most Cariban languages having only between 100 and 3,000 speakers. Carib proper (its English name, called Galibi, Kaliña, Kari'na, and Cariña in other countries) is spoken by between 10,000 and 25,000 people along a roughly 1,000-mile arc of the coast and inland from the northeasternmost point of Brazil through French Guiana, Surinam, and Guyana, to the easternmost 300 miles of the Venezuelan coast. Several Caribbean languages historically claimed to be Cariban (Black Carib, Island Carib, Garifuna) are linguistically Arawakan, with some Carib features (mostly vocabulary) due to intensive contact with Carib invaders.

Most languages of the Cariban family are still poorly documented, with unreliably transcribed word lists and poor or no grammatical description. The two best documented Cariban languages are Carib (due mostly to work by Berend Hoff) and Hixkaryana (due exclusively to work by Desmond Derbyshire); work by Sergio Meira on Tiriyó is rapidly bringing a third language into this select group. The Carib language boasts of two grammars, two dictionaries, a collection of texts, several articles on selected linguistic topics and work in progress in both French Guiana and Venezuela; Hixkaryana has two grammars, a collection of texts and several articles, all by Derbyshire. Other Cariban languages with grammars (of uneven quality and comprehensiveness) include Apalaí, De'kwana, Ikpéng (Txikao), Makushi, Pemón, and Waiwai. In the domain of the lexicon, Cariban languages have been represented almost exclusively by bilingual word lists (some labeled dictionaries), with true dictionaries existing for only Carib, Panare, and Pemón. Scholars and students are actively working on projects in every country where Cariban languages are spoken. Ongoing research projects that ought to result in solid primary descriptive materials (text collections, grammars, dictionaries) include: in Brazil, Arara, Ikpéng (Txikão), Ingarikó (Kapóng), Katxúyana, Kuikúru, Tiriyó, Waimiri-Atroari, and Wayana; in Colombia, Yukpa; in French Guiana, Wayana and Kali'na; in Guyana, Akawaio (Kapóng); in Surinam, Tiriyó; and in Venezuela, Japrería, Kari'ña, Mapoyo/Yawarana, Panare, Pemón, and Yukpa.

Given the poor documentation of most Cariban languages, current classifications of the Cariban family are all unconvincing, as are reconstructions of ancestral homelands and migration patterns. The most recent classificatory hypothesis (which will certainly evolve as more data become available) is reproduced here in Table 1; it is derived by modifying Terrence Kaufman's (1994) classification (i) adding suggestions from recent work in Venezuela and the Guiana Plateau, (ii) marking as extinct (†) every language that has not been attested in the last 25 years, (iii) using current

TABLE 1

Venezuelan Branch (A–B–C–D–E–F)

Pemóng–Panare–Mapoyo–Tamanaku Sub–Branch (A–B–C–D)

Pemóng–Panare Macro-Group (A–B)

A. Pemóng Group

  A1. Pemóng Proper Subgroup: Makushi, Pemóng (Taurepang, Kamarakóto, Arekuna), Kapóng (Akawaio, Patamuna, Ingarikó)

  A2. (†)Purukotó

B. Panare

Mapoyo-Tamanaku Macro-Group (C–D)

C. Mapoyo (Mapoyo, Wanai, Yawarana, Pémono)

D. †Tamanaku

E. †Kumaná (Chaima, Cumanagota)

F. Makiritare (De'kwana, Ye'kwana, Maiongong)

Southern Branch (G–H–I)

G. Nahukwa Group: Kuikúru, Kalapalo, Amonap, Matipú

H. Bakarí

I. Arara Group: Arara-Pirirí, Ikpéng (Txikão), †Apiaká-Apingi, †Juma, †Yarumá

North Amazonian Branch (J–K)

J. Jawaperi Group: Waimirí-Atroarí, †Jawaperi, †Bonari

K. Paravilyana Group

  K1. †Sapara

  K2. Paravilyana Subgroup: †Pawishiana, †Paravilyana

Residue (groups and Languages still in search of branches, in alphabetical order)

Groups

L. Parukotoan Group

  L1. Katxúyana (Chikena, Cachuena, Ingarüna, Shikuyana, Warikyana)

  L2. Waiwai SubGroup: Waiwai (Wabui), Hixkaryana

M. Taranoan Group

  M1. Tiriyo Subgroup: Akuriyo, Tiriyo (Trio, Saluma, Pianakoto)

  M2. Karihona (Carijona, Hianácoto)

N. Wayana Group: Wayana, †Arakuajú

O. †Yao Group: †Tiverikoto, †Yao

P. Yukpa Group: Yukpa, Japreria, †Koyama

Languages

Q. Apalaí

R. Carib (Kari'nya, Kalinya, Cariña, Galibi)

S. †Opon-Karare

T. †Palmella

U. †Pimenteira

local names for each language, and (iv) conservatively isolating several languages/groups that Kaufman had linked into tentative branches.

In the absence of a solid linguistic classification, discussions of prehistory — especially ancestral homelands and migration patterns — are perforce somewhat speculative. One hypothesis suggests that the ancestral homeland of Proto-Carib was in the Guiana plateau, from which the outlying groups moved southward and westward. Another posits a central Brazilian origin with one or more subsequent northward migrations and rapid expansion upon arrival in the Guiana Plateau. In either scenario, Cariban peoples appear to have occupied the Guiana Plateau for at least three millennia.

Characteristics of Carib and Cariban languages

Phonology
The Carib language has a relatively straightforward sound system, with three voiceless stops /p, t, k/, a voiceless fricative /s/, two nasals /m, n/, two glides /w, j/, and a sound that—depending on the neighboring sounds—surfaces strictly in syllable-final position as either a velar fricative /x/ or a glottal stop /ʔ/. All consonants are palatalized (with tongue movement toward the palate) when adjacent to [i]; /p, t, k/ are pronounced as /b, d, g/ between vowels or after nasals and glides; and the pronounciation of nasals depends on the following consonant. Some Cariban languages (and even some Carib dialects) add further consonants to the above inventory.

The six-vowel system of Carib (and most other Cariban languages) includes three high vowels /i, ɨ, u/, two mid-vowels /e, o/ (usually phonetically [ɛ, ɔ]), and one low vowel /a/. Across the family, the most common addition to this inventory is /ə/, with a few languages adding a second set of mid-vowels [ɛ, ɔ]. Independently, nasalized vowels are attested only in Apalaí. The difference between long and short vowels length is a differentiating factor for words in Carib and several other Cariban languages, but vowels may also be lengthened due to speech rhythm: in an iambic foot structure, the vowel of every second light (short, open) syllable is lengthened (excepting the final syllable, which is never lengthened). Across the family, primary intonational stress has been difficult to identify reliably, and sometimes it appears not to fall on the head of a foot (e.g. in Carib, Hixkaryana, Pemón, and Tiriyó).

Morphology
Carib makes heavy use of affixes, i.e. word particles are added to other words to either add to or change the meaning. In some cases, words together with their affixes may express meanings equivalent to phrases or sentences in other languages. Personal prefixes ('I', 'you', 'he'/'she'/'it') and grammatical prefixes or suffixes appear on verbs, nouns, and postpositions. Carib has suffixes that change nouns to verbs, nouns or verbs to adverbs, and adverbs, verbs, or postpositions to nouns. Most Cariban languages match this profile, with changes usually in the direction of simplification. There are two word classes in Cariban languages, particles and sound symbolic words, to which no affixes may be attached; both are rare in elicited speech but ubiquitous in texts.

Syntax
In Carib and most languages across the family, certain two-word combinations—possessor–possessed, object–postposition, and object–verb—are clearly grammatical units, the tight bond between them demonstrated by invariant order, the inability to interpose words between them, and the fact that they seem to be pronounced as a single word. Beyond these two-word sequences, the word order appears to be quite variable, in Carib or in other Cariban languages.

The Carib main clause consists of an inflected verb, with affixes indicating the type of both subject and object. Noun phrases can optionally occur as subject and object. When both subject and object of a transitive verb are third person, the object immediately precedes the verb, forming a tight verb phrase in a subject–object–verb construction. The subject usually precedes the verb phrase, but this is not always required. When either subject or object is not third person, the subject–object–verb ordering (SOV) remains more frequent, yet all other orders of subject, object, and verb are permitted. Carib distinguishes three degrees of past tense, and each tense has both a perfective and an imperfective form. A similar syntax is found in 19 other Cariban languages, belonging to virtually every proposed branch of the family, which indicates that the common ancestral language (Proto-Carib) also had this structure. Several Cariban languages present the rare basic order OVS, documented first — and most extensively — in Hixkaryana. The OVS order is commonly observed also in languages where it is not clearly 'basic'. While SVO is attested in several Cariban languages, the predominance of SVO in a few languages is most likely attributable to contact with languages of colonization.

Ergativity is pervasive in the Cariban family, and indeed throughout much of South America. In ergative constructions, the subject of an intransitive verb and the direct object of a transitive verb (called the absolutive) receive the same grammatical treatment, that is, they receive the same case-marking, verb agreement, word order, etc. In contrast, the subject of a transitive verb (called the ergative) takes different case-marking, verb agreement, word order, etc. Languages across the family do not use relative clauses, i.e. subordinate clauses that are introduced by words like *that, which, who,* etc. Instead, Cariban languages use affixes to turn the subordinate verb into a noun or an adverb.

Early hypotheses about historical evolution of Cariban grammar held that the ergativity seen in subordinate clauses indicates that Proto-Carib had ergative syntax in both main and subordinate clauses, with the nonergative main clause in today's languages being an innovation. Most recent work argues that the few cases of main clause ergativity in the Cariban family are readily explained as innovations, in which ergatively organized subordinate clause morphosyntax is 'elevated' to main clause status. The origins of all main clause types observed so far in the northern Cariban languages fit this latter hypothesis; however, each new description, especially from the less-known southern Cariban languages of the Xingu valley, will provide new challenges to the historical reconstructions in place so far.

References

Derbyshire, Desmond C. 1985. Hixkaryana and linguistic typology. Arlington: Summer Institute of Linguistics and the University of Texas at Arlington.

Derbyshire, Desmond C. 1999. The Cariban language family. Languages of South America, ed. by R.M.W. Dixon and Alexandra Aikhenvald. Cambridge: University of Cambridge Press.

Gildea, Spike. 1998. On reconstructing grammar: comparative Cariban morphosyntax. Oxford: Oxford University Press.

Hoff, Berend J. 1968. The Carib language. The Hague: Martinus Nijhoff.

Kaufman, Terrence K. 1994. The Native languages of South America. Atlas of the world's languages, ed. by Christopher Mosely and R.E. Asher. New York: Routledge.

SPIKE GILDEA

Caribbean

The Caribbean islands (collectively called the West Indies) extend over a thousand miles from the tip of the Florida Peninsula to the northern coast of South America. They vary in size and population, ranging from Cuba, with an area of 42,804 square miles and more than ten million inhabitants, to the small, uninhabited, rocky islets of the Grenadines and the Virgin Islands. Most islands are smaller than Barbados (only 166 square miles), with a quarter of a million inhabitants. Although the islands share in common a long history of colonialism beginning in the fifteenth century, slavery, and economies once dominated exclusively by sugar plantations, each island has its distinctive history, culture, and language or languages. Politically, they comprise 23 entities, ranging from those that are still quasicolonies, such as the French Antilles (a French *département*) and Puerto Rico (officially a commonwealth of the United States), to those that are independent, such as Haiti (formerly a French colony), the Dominican Republic (originally a Spanish colony), and Jamaica (formerly a British colony). Some territories, such as Guyana, have passed through a series of colonial regimes before becoming independent.

Divided linguistically and culturally, the West Indies can be categorized into four distinct groups according to their officially recognized languages, usually the language of the former or current colonizer. The largest group is the Anglophone Caribbean, including the majority of islands outside the Greater Antilles (i.e. Anguilla, Antigua and Barbados, the Bahamas, Barbados, British Virgin Islands, Cayman Islands, Dominica, Grenada, Jamaica, Montserrat, St. Kitts-Nevis, St. Lucia, St. Vincent and the Grenadines, Trinidad and Tobago, Turks and Caicos islands, and the United States Virgin Islands), as well as two mainland nations: Belize in Central America and Guyana in South America. The Hispanophone Caribbean includes Puerto Rico, Cuba, and the Dominican Republic. The Francophone Caribbean includes Haiti and the French Antilles (Guadeloupe, Dominica, Martinique, and St. Lucia) in the Lesser Antilles island group. The Netherland Antilles (Aruba, Bonaire, Curaçao, St. Eustatius, Saba, and St. Maarten) are all that remain of the former Dutch empire in the Caribbean.

The native American population was exterminated first by the Spanish in the Greater Antilles (Cuba, Puerto Rico, and the Dominican Republic) and by the British and French in the Lesser Antilles. The indigenous Arawakan languages became extinct as their speakers died as a result of conquest, disease, or forced labor, and their remnants were absorbed in the Hispanicized population. The islands were repopulated through the introduction of African slaves to cultivate sugar and coffee plantations established by European settlers during the seventeenth and eighteenth centuries. After the abolition of slavery in the British, French, and Dutch territories in the nineteenth century, indentured labor was

introduced for plantation work, particularly from northern India.

The historical partition of the West Indies among the various European colonizers (Spanish, French, British, Dutch, and Danish) has had a number of important consequences. Eric Williams, first and long-time Prime Minister of newly independent Trinidad and Tobago (1962–1981) and founder of the People's National Movement, which led the country to independence, believed that the absence of an indigenous Caribbean language was a major obstacle to decolonization. This meant that there was no rallying point for the nationalist movement. Moreover, the absence of a common language was a hindrance to political, economic, and cultural contacts between the islands. The communication gap existing between the elite, who speak metropolitan European languages, and the masses, who speak creoles, also poses serious problems of social and political organization. Interisland connections have remained weak, because each colonizing power tended to dominate relationships with its own colonies. Where interisland linkage has occurred, it has been confined largely to islands belonging to the same European language group (each of these usually being of the same colonial affiliation). The Hispanophone islands also have a link to Latin America, which is lacking in the English-, French-, and Dutch-speaking territories. Although plantations continue to occupy large portions of land, their social significance has declined as people have left the countryside, either moving to the towns or emigrating to other countries.

Major metropolitan languages of European origin that play a role in the Caribbean today are the languages of the former and present colonial administrations: English, Dutch, French, and Spanish. In addition to the standard varieties of these languages, there are a number of colonial varieties of French in St. Barthélemy (St. Barts) and St. Thomas. Standard Dutch is an official language in the Netherlands Antilles, the only remnant of the Dutch empire. The Caribbean basin constitutes a clearly definable dialect area of American Spanish, including the varieties spoken in Cuba, Puerto Rico, the Dominican Republic, a few places in Trinidad, Guyana, the Caribbean coast of Venezuela and Colombia, and some coastal areas of Central America and Mexico.

Languages of Southeast Asian and Asian origin include varieties of Hindi, Urdu, and Chinese brought by indentured laborers. These languages are primarily nonstandard varieties used as home languages with no wider currency or status, and the younger generation are often no longer fluent in their use.

In their everyday lives, the majority of people in the Caribbean speak creole languages, which have developed from contact between the languages of the colonizers and those of the slave laborers. Nevertheless, the majority of children throughout the Caribbean continue to be educated in metropolitan languages rather than in their own native creole languages. The disparity between home and school language has contributed to a high rate of educational failure and low rates of literacy. Compare Haitian Creole French *Mwen grangou* with standard French *J'ai faim* 'I'm hungry'. Public attitudes to the creoles still tend to be largely negative, and the comments made by some speakers of French Creole from the Eastern Caribbean are typical: 'It's broken French, you can't write it down. No, it's not a language'. Speakers often have no special names for their languages other than *patois* (patwa) or broken English/French.

Despite the school's rejection of the children's native language, creoles today are the most powerful vehicles by means of which local Caribbean identities are formed and expressed. The absence of a high culture expressed through an indigenous language has meant that the voice of the masses and popular culture has generally been expressed through creoles, in both spoken and written forms.

In Jamaica, in particular, a strong, black national identity has gone hand in hand with nationalist politics to foster a powerfully vibrant creative arts movement, whose influence has extended far beyond Jamaica and the Caribbean. Rejecting the term *dialect* because it suggested inferiority, Edward Kamau Brathwaite (1984) argued for the use of what he called 'nation language' (Jamaican Creole English) in poetry as a way of capturing the sounds and rhythm of oral traditions of performance. Inspired by hearing a recording of T.S. Eliot reading from *The Waste Land*, Brathwaite urged poets to model their poetry on the African-derived rhythms of calypso to break the pentameter, which other New World poets before him, such as Walt Whitman, had also sought to undermine. Brathwaite was one of the founders of the Caribbean Artists Movement in 1966 in London, who fostered the development of a West Indian literature rooted in the languages and experiences of the islands. Through the commercial success of performers such as Bob Marley, Mikey Smith, Linton Kwesi Johnson, and Benjamin Zepphaniah, or Mutabaruka, whose music and sound poems were on the British reggae music charts in the late 1970s and early 1980s, the once historically devalued Caribbean popular culture has become part of multicultural Britain.

The Anglophone Caribbean has a population of approximately six million. The three largest

English-creole-speaking communities are Jamaica, Trinidad and Tobago, and Guyana. In most of these communities, standard (usually British) English is the official language and is used in most public sectors of communication. Despite the fact that the creoles of the eastern Caribbean are older, they are closer to standard English than the creoles of the western Caribbean. This is because of the presence of fewer Africans in the early settler population, allowing regional forms of English to be established.

In a small number of places, the creoles coexist with a French creole (e.g. in St. Lucia and Dominica) or with other ethnic languages (e.g. Belize, Guyana, United States Virgin Islands, and Trinidad). In most cases, there is a broad spectrum of varieties, which can be arranged along a continuum from those that are 'deepest', i.e. furthest removed from standard English and spoken by the rural uneducated, to those that are closest to standard English. Most people know several varieties and are able to switch between them depending on the context. Compare some of the variants from Guyanese Creole, *mi gii am/mi bin gii am/mi bin gii ii* with *a geev ii/a geev him* 'I gave him'. The deepest form of creole uses the object form *me* instead of standard English *I*; past tense marking is optional by means of *bin*.

The Afro-Caribbean region has been one of the vibrant areas in the world producing literature in English, as evidenced by the fact that Nobel prizes and other prestigious awards have gone to some of its writers, such as Wole Soyinka from Nigeria and Derek Walcott (born 1930) from St. Lucia, some of whose works draw on pidgin and creole English.

French creoles are spoken primarily on the French Antilles islands (Guadeloupe, Dominica, Martinique, and St. Lucia) of the Lesser Antilles and Haiti, as well as in former French Guaiana on the South American mainland between Suriname and Brazil, today a département of France. Although each creole variety spoken in the French Antilles is slightly different, they are sufficiently similar to one another to constitute a group distinct from Haitian and from the varieties spoken in the Commonwealth Antilles (Trinidad and the Windward Islands of Dominica, St. Lucia, and Grenada), which were British from the eighteenth century until their recent independence. In these countries, English is now the official language. Hence, the creoles coexist with English creoles and standard English and have been influenced by English, whereas Haitian and the French Antilles varieties have continued to be influenced by French. Creole French is still the main vernacular on Dominica and St. Lucia, although English and creole English are gaining ground. There is a fairly strong tradition of written literature, particularly poetry, in the French Antilles. Haiti, occupying the western third of the island of Hispaniola, which it shares with the Dominican Republic, has the largest number of creole speakers (more than six million) in the Caribbean. Independent since 1984, 90% of Haitians speak creole as their only language, and are culturally and linguistically isolated from France. A bilingual elite has continued to maintain French, and French continues to dominate in the public and formal sectors, such as government and education. Since 1983, Haitian Creole has been a co-official language with French and has been used in some primary schools as well as in adult literacy campaigns. Haitian Creole has one of the richest literatures of any creole languages, dating back to the nineteenth century. A creole variety of Spanish called Palenquero is spoken by older people in an isolated village called El Palenque de San Basilio, south of the city of Cartagena on the Caribbean coast of Colombia. Papiamentu is a creole based on Portuguese and Spanish spoken on the Leeward Islands of the Netherland Antilles (Aruba, Bonaire, and Curaçao) just north of Venezuela. Papiamentu has the highest status of any of the Caribbean creoles; it is spoken by virtually all social classes across a wide range of contexts and is used in the media.

Only three creole forms of Dutch are known to have resulted from the presence of Dutch traders and settlers in the Caribbean during the seventeenth and eighteenth centuries. A variety called Negerhollands ('Black Dutch') was present until just recently in the Virgin Islands, ruled first by Denmark and then by the United States. There it was replaced by local forms of English. The other two nearly extinct varieties, Berbice Creole Dutch and Skepi Creole Dutch, arose in what were once separate colonies isolated by great rivers and wildernesses, but became subsequently British Guiana (later independent Guyana) in northern South America. They are now spoken only by older people and are not mutually intelligible.

References

Allsopp, Richard. 1996. Dictionary of Caribbean usage. Oxford: Oxford University Press.

Brathwaite, Edward Kamau. 1984. The history of the voice: The development of nation language in Anglophone Caribbean poetry. London: New Beacon Books.

Dalphinis, Morgan. 1985. Caribbean & African languages. London: Karia Press.

Devonish, Hubert. 1986. Language and liberation: Creole language politics in the Caribbean. London: Karia Press.

Holm, John. 1989. Pidgins & Creoles. Vol. II. Reference survey. Cambridge: Cambridge University Press.

Holm, John. 1994. English in the Caribbean. Cambridge history of the English language, Vol. V, English in Britain and overseas: origins and development, ed. by Robert Burchfield. Cambridge: Cambridge University Press.

Lawton, David L. 1982. English in the Caribbean. English as a world language, ed. by Richard W. Bailey and Manfred Görlach. Ann Arbor: University of Michigan Press.

Morris, Carol and Hubisi Nwenmely. 1993. The Kwéyôl language and literacy project. Language and Education 7(4). 259–70.

Ramchand, Kenneth. 1970. The West Indian novel and its background. London: Faber.

Williams, Eric. 1970. From Columbus to Castro: the history of the Caribbean 1492–1969. New York: Harper & Row.

Winford, Donald. 1991. The Caribbean. English around the world. Sociolinguistic perspectives, ed. by Jenny Cheshire. Cambridge: Cambridge University Press.

SUZANNE ROMAINE

See also **English; French Language; Pidgins and Creoles**

Case

The notion of case was inherited from traditional grammars describing languages such as Latin, Sanskrit, Russian, and German. In these languages, nouns (or noun phrases) differ according to their particular function in the sentence; i.e. a subject noun would be slightly different when used as an object. Languages differ very much with respect to how case is manifested. There are languages, such as Chinese, that do not mark nouns for case at all. There are also languages with a partial manifestation of case. In English, for example, regular noun phrases do not show case marking, but pronouns do: *he* (nominative case, subject) vs. *him* (accusative case, object) vs. *his* (genitive, or possessive, case). Then, there are languages that have a rich case system, such as German, in which regular noun phrases are marked for case: *Der Hund bellt* 'The dog barks' (nominative, subject) vs. *der Kopf des Hundes* 'the dog's head' (genitive) vs. *Ich gebe dem Hund einen Ball* 'I give a ball to the dog' (dative, indirect object) vs. *Ich sehe den Hund* 'I see the dog' (accusative, direct object).

Given such a variation, there was a long debate in linguistics on whether case is universally active in all languages or whether this notion is relevant only in the description of languages that actually mark their nouns accordingly. In the first half of the twentieth century, the most influential theories on case ignored its syntactic nature and aimed to specify the meaning of different case markers (prepositions, inflections, etc.). Charles Fillmore (1968) introduced the notion of case into the framework of generative grammar, arguing that case reflects meaning relationships between e.g. the verb and the subject or object in a sentence. In the late 1970s, however, Jean-Roger Vergnaud shifted perspective and claimed that case is a manifestation of syntactic relations only. In other words, case marking shows only whether or not a given noun is e.g. the subject or an object, independent of the meaning of the words involved. Noam Chomsky's *Lectures on Government and Binding* (1981) gives a first explicit formulation of case theory as a separate module of Universal Grammar.

The Chomskyan theory of case relies crucially on the assumption that case is present in all languages (in Chinese as well as in English or Russian) and that overt case marking is just a manifestation of an abstract case, which is assigned to noun phrases in a particular syntactic environment. Although traditional grammars use a fairly large number of cases, Chomsky eliminated all but three: nominative (assigned to the subject position of a finite clause, e.g. *John/he speaks well*), accusative (assigned to an object position of a transitive construction, e.g. *We see John/him*), and oblique (assigned by prepositions, e.g. *of John/him, to John/him,* etc.).

For case to be functional, the syntactic environment has to fulfill certain conditions, among them the requirement that a 'case assigner' has to be present. In English, verbs, prepositions, and tense can function as case assigners. A sentence such as *I am happy John* is ungrammatical, because there is no word adjacent to *John* that could possibly assign oblique case to it. *I am happy for John*, on the other hand, does contain the case-assigning preposition *for* and is thus grammatical.

Furthermore, a noun phrase must stand in a particular relation with its case assigner. A sentence such as *John believes that him to be happy* is ungrammatical, because the noun *him* is structurally too far from the case-assigning verb *believes* to receive the proper

accusative case. However, the verb is able to exert the proper influence on the noun phrase if the complementizer *that* is absent: *John <u>believes him</u> to be happy.*

Thus, the case assigner and the case-receiving noun phrase have to be directly adjacent. This is also shown by the contrast between the following sentences: *I see rarely him* and *I see him rarely*. The first one is ungrammatical, because the adverb *rarely* intervenes between the noun phrase *him* and the case-assigning verb *see*. The second one shows *rarely* at the sentence-final position, *see* is adjacent to *him*, and accusative case can be assigned. The sentence is thus grammatical.

According to early syntactic theories, case was discussed solely with respect to the form of nouns. However, Chomsky argued in *Knowledge of Language* (1986) that case seems to be crucial for the proper interpretation of the sentence. Thus, case is not only a formal requirement but also plays a role in meaning.

Case is an abstract notion and can be divided into two kinds: 'structural case' and 'inherent case'. Nominative and accusative are considered structural cases, because they refer only to the formal properties of the sentence: they mark whether a given noun is a subject or an object. Dative, genitive, etc., i.e. the oblique cases, are inherent cases because they depend on a meaning relationship between the case assigner and the noun phrase that is assigned the case.

The notion of inherent case implies that nouns, as well as adjectives, can be regarded as case assigners. Thus, in *a gift <u>to</u> my friend* and *a gift <u>of</u> my friend,* the case relationship is one between the two noun phrases *a gift* and *my friend*. Thus, *a gift* is said to be the case assigner, whereas *my friend* receives case. The prepositions *to* and *of* are interpreted as realizations of the dative and genitive cases, respectively.

'Abstract case' remains a very important concept in linguistics. The question about the nature of case and its role in syntax is still open.

References

Blake, Barry J. 1994. Case. Cambridge: Cambridge University Press, 2nd edition, 2001.

Chomsky, Noam. 1981. Lectures on government and binding. Dordrecht: Foris.

Chomsky, Noam. 1986. Knowledge of language: its nature, origin, and use. New York: Praeger.

Fillmore, Charles. 1968. The case for case. Universals in linguistic theory, ed. by Emmon Bach and Robert T. Harms. New York: Holt, Rinehart, and Winston.

Haegeman, Liliane. 1991. Introduction to government and binding theory. Oxford and Cambridge, MA: Blackwell; 2nd edition, 1994.

Lasnik, Howard. 1999. Minimalist analysis. Oxford and Malden, MA: Blackwell.

EGOR TSEDRYK

See also **Chinese (Mandarin); Chomsky, Noam; English; German; Grammar, Traditional; Universal Grammar**

Causation

The notion of CAUSATION, whereby an agent causes an event to occur has long been considered to be a fundamental semantic primitive. This can be seen in work in a variety of traditions: Generative Semantics, Functional Linguistics, Cognitive Linguistics, Conceptual/Event Structure, Role and Reference Grammar, and Proto-Roles, to name a few. Many of these approaches assume that there is a fundamental relation between an agent of causation (CAUSER) and the CAUSED EVENT. In some cases, the causation is effect through an intermediary (CAUSEE).

The expression of causation can be intrinsic to certain predicates (LEXICAL CAUSATIVES) or can be part of a productive process by which a language encodes causation (PRODUCTIVE CAUSATIVES). Lexical causatives include predicates that inherently denote a causative relation (1) or can form part of an inchoative/causative alternation (2). Productive causatives involve a separate causative predicate, either a separate word (ANALYTIC CAUSATIVES—3a) or a causative morpheme (SYNTHETIC CAUSATIVES—3b):

Lexical Causatives:

(1) The farmer killed the duckling. *Inherent causative*

(2) a. The window broke. *Inchoative*
    b. The child broke the glass. *Causative*

Productive Causatives:

(3) a. The farmer made the duckling die. *Analytic causative*

    b. Adam çocuğa kaıyı aç-**tür**-dı *Synthetic causative (Turkish)*

    'The man made the child open the door.'

A good deal of interest with respect to the linguistics of causation comes from the manners in which languages tend to productively encode the notion of causation. This article surveys two dimensions of the encoding of causation: CLAUSALITY and CAUSEE ENCODING.

A striking aspect of productive causative constructions is the manner in which they simultaneously exhibit both biclausal and monoclausal characteristics. Because productive analytic causatives have two distinct predicates, a base predicate (e.g. *die* in 3a) and a causative predicate (e.g. *make*), it is not surprising that they behave as two clauses. More surprising are the monoclausal aspects of these constructions. For example, analytic causatives in Spanish appear to be standard biclausal constructions with embedded infinitival clauses:

(4) El maestro les hizo leer ese libro a los estudiantes.

    'The teacher made the students read that book.'

However, on closer inspection, various local dependencies, which are normally restricted to single clauses, can span the two predicates. For example, the placement of object clitics is normally clause-bounded; the clitic occurs with the verb of which it is an object. However, in the case of causative constructions, a clitic may (and sometimes must) occur with the causative predicate (known as CLITIC CLIMBING):

(5) El maestro se **lo** hizo leer a los estudiantes.

    'The teacher made the students read it.'

This is one of several clause-bounded phenomena that are mitigated across causative and base predicates, and a way in which causative constructions, even when expressed analytically, share characteristics with monoclauses (cf. Aissen 1979; Aissen and Perlmutter 1983; Rizzi 1978; Burzio 1986, among others). This has led to analyses in terms of COMPLEX PREDICATE formation. While the mechanics of such analyses have differed, the basic insight is the same: two predicates define a single clausal domain (see Moore (1996) for a survey of approaches).

Because the causative and base predicates form something like a single, complex predicate, the arguments of the two predicates become coarguments. In this way, the argument that would have been the subject of the base predicate (the CAUSEE argument) can no longer be a subject, as that role is taken by the CAUSER argument. While languages vary in how the causee argument is encoded, there is a good deal of cross-linguistic consistency; this is true regardless of whether a productive causative is expressed analytically or synthetically (see Gibson and Raposo 1986). Two basic patterns are widespread—one having to do with the valence of the base predicate, and the other with the semantics of the causative predicate:

(6) *Causee encoding—valence based*:

    (a) If the base predicate is INTRANSITIVE, the causee is encoded as a DIRECT OBJECT.

    (b) If the base predicate is TRANSITIVE, the causee is encoded as an INDIRECT OBJECT or as an OBLIQUE.

(7) *Causee encoding—semantically based:*

    (a) If the causee undergoes DIRECT CAUSATION, it is encoded as a DIRECT OBJECT.

    (b) If the causee undergoes INDIRECT CAUSATION, it is encoded as an INDIRECT OBJECT or as an OBLIQUE.

Spanish causatives exhibit both of these encoding patterns:

(8) *Causee encoding based on valence of base predicate:*

    a. José *los* hizo trabajar. *Intransitive base predicate*

    'Jose made *them*$_{DO}$ work.'

    b José *les* hizo leer el libro.
Transitive base predicate

    'Jose made *them*$_{IO}$ read the book.'

(9) *Causee encoding based on directness of causation:*

    a. *La* hice probarlo a la fuerza.
Direct causation
'I made *her*$_{DO}$ try/taste it by force.'

    b. *Le* hice probarlo diciéndole que era riquísimo.
Indirect causation
'I had *her*$_{IO}$ try/taste it by telling her it was delicious.'

Analyses that deal with the two types of encoding patterns include those of Comrie (1981), Cole (1983), and Ackerman and Moore (1999).

Causation is a pervasive semantic notion that exhibits strikingly consistent patterns of grammaticization cross-linguistically.

References

Ackerman Farrell, and John Moore. 1999. Syntagmatic and paradigmatic dimensions of causee encodings. Linguistics and Philosophy 22. 1–44.

Aissen, Judith L. 1979. The syntax of causative constructions. New York: Garland.

Aissen, Judith L., and David M. Perlmutter. 1983. Clause reduction in Spanish. Studies in relational grammar, Vol. 1, ed. by David M. Perlmutter. Chicago: The University of Chicago Press.

Burzio, Luigi. 1986. Italian syntax: a government and binding approach. Dordrecht: Reidel.

Cole, Peter. 1983. The grammatical role of the causee in universal grammar. International Journal of American Linguistics 49. 115–33.

Comrie, Bernard. 1981. Language universals and linguistic typology: syntax and morphology. Chicago: The University of Chicago Press.

Gibson, Jeanne, and Eduardo Raposo. 1986. Clause union, the stratal uniqueness law, and the Chômeur relation. Natural Language and Linguistic Theory 4. 295–331.

Jackendoff, Ray. 1991. Semantic structures. Cambridge, MA: MIT Press.

Moore, John. 1996. Reduced constructions in Spanish. New York: Garland.

Rizzi, Lugi. 1978. A restructuring rule in Italian syntax. Recent transformational studies in European languages, ed. by Samuel J. Keyser. New York: Academic Press.

Strozer, Judith Reina. 1976. Clitics in Spanish. Doctoral Dissertation. Los Angeles: University of California.

Talmy, Leonard. 1985. Force dynamics in language and thought. Papers from the parasession on causatives and sgentivity, ed. by William H. Eilfort, Paul D. Kroeber, and Karen L. Peterson. Chicago: Chicago Linguistic Society.

JOHN MOORE

Celtic Languages

'*Gallia est omnis divisa in partes tres, quarum unam incolunt Belgae, aliam Aquitani, tertiam qui ipsorum linguae Celtae, nostra Galli appellantur. Hi omnes lingua, institutis, legibus se differunt.*'

'All Gaul is divided into three parts, one of which the Belgae inhabit, the Aquitani another, those who in their own language are called Celts, in our Gauls, the third. All these differ from each other in language, customs and laws.' (Caesar, *De bello gallico* I,1)

Celtic is classified as a separate branch of the Indo-European (IE) language family. Like Greek and Latin, Celtic is a Centum language. It can be further subdivided into Continental and Insular Celtic. The division into Continental and Insular Celtic languages is not merely geographically based, but also genetically based. While the Gaulish and Brittonic languages share some features, excluding Celtiberian and Goidelic, the Insular and Continental division is based on commonalties in sound shifts from IE.

Although scholars agree on the relationship between ancient and modern Celtic languages, it must be stressed that speakers of different Continental Celtic languages did not refer to themselves as mutually related 'Celts'. These languages died out before the end of the Roman Empire, while evidence of the Celtic languages in the British Isles starts in the Early Middle Ages. This makes comparison of ancient modern Celtic languages problematic. The living Celtic languages can be further subdivided into Goidelic (Gaelic) and Brittonic (British). The descendants of Goidelic are Modern Irish and Scottish Gaelic spoken in Ireland and Scotland and the now extinct Manx, whose alleged last native speaker died in 1974. The Brittonic Celtic languages are Welsh in Wales, Breton in Britanny and Cornish, which died out in the eighteenth or nineteenth century.

In pre-Roman times, Continental Celtic languages were spoken all over Europe. Lepontic is attested in northern Italy before the arrival of the Gauls. Gaulish (in today's France) is known through personal names in Latin texts and some inscriptions in Greek script and a considerable amount in Latin; Celtiberian (or Hispano-Celtic) on the Iberian peninsula in Iberian script (Iberian being a non-Indo-European language), Galatian in Asia Minor, today's Turkey. A contested issue is the 'celticity' of Lusitanian, attested in three pre-Roman inscriptions in the west of the Iberian peninsula. Archeologically, the original native place of the Celts must have been in Central Europe, somewhere between Bohemia and Bavaria. The earliest clearly datable sources go back to the sixth century BCE.

Today, the surviving Celtic languages are confined to the westernmost parts of Europe, the *Celtic fringe*: the west, northwest and southwest of Ireland, Scottish Highlands and Western Isles, large parts of Wales and western Britanny. Remnants of Celtic-speaking people are also to be found in the Americas: Welsh-speaking people in Patagonia (Argentina) and some Scottish Gaelic in Nova Scotia (Canada). Exactly when and in how many migration waves the Celts came to Ireland is hard to determine. The oldest Irish coherent manuscript text *Amra Choluim Cille* goes back to 600 CE,

making Irish the oldest attested continental European language after Greek and Latin. Within the written Irish tradition, three main periods can be distinguished: Old Irish, Middle Irish, and Modern Irish. The Great Famine (1845–1948) nearly eradicated the Modern Irish language. More than one million people starved, while another million fled to America. Today, after efforts to revive the endangered language, only three main dialects are distinguished according to the provinces in which they are spoken: *Connacht, Ulster,* and *Munster Irish. A* standard language, mainly based on Munster Irish, also exists. Irish has been the official language of Ireland since the establishment of the Free State in 1922.

In the study of Celtic languages, phonology plays a large role as Old Irish preserves few case or personal endings. As early as Common Celtic times, the sound *p* had disappeared. Therefore, the IE word for *father* appears in Old Irish as *athair.* According to some scholars (see Untermann and Wodtko 1997), the existence of *p* in Lusitan makes it probable that this loss was not a common Celtic phenomenon. Another peculiarity of Celtic languages are the so-called initial mutations. Simply speaking, mutation means that the sound dropped at the end of one word affects the first sound of the following word. One of the most interesting characteristics of Celtic languages is their syntax: the basic word order is Verb–Subject–Object, where the verb stands at the beginning of the sentence. For example, in Irish, *D'oscail mé an doras* (lit. 'Opened I the door'), the verb not the subject is first.

Irish

The sound system of Irish is fairly complicated compared to English. For each consonant, there is a palatal equivalent, making it similar to Russian. Lenition, a sound assimilation, and eclipsis, the omission of symbols, make the orthography look somewhat strange to the native speaker of English. Examples include phrases like *i nGaillimh* and *ó Bhaile Átha Cliath* for 'in Galway' and 'from Dublin' and surnames like *Ní Dhomhnaill* for female 'Mac Donald'. Two verbs for 'to be' exist in Irish, one being the copula *is* and the other the substantive verb *tá.* Their distribution is very similar to Spanish *ser* and *estar.* There is no word for 'to have' in Irish, as for many—in other languages clearly verbal concepts like 'to love', 'to owe', 'to know', etc., are all phrasal verbs containing the verb 'to be' (e.g. Ir. *Tá grá agam duit* lit, 'there is love at me to you'). This gives us a picture of a language that on the one hand has a clearly verb-centered syntax, but on the other very few strictly verbal concepts.

Scottish Gaelic

Before the Middle Ages (approximately tenth century CE), there was no difference between the Gaelic spoken in Scotland and the one in Ireland, since they both developed from the same parent language, Old Irish. The Reformation played a key role in linguistic self-definition. After the Jacobite Rising in 1745, a huge deportation of Scottish Gaelic-speaking population to Overseas initiated the blackest chapter in the history of Scotland: the Clearances. In the first half of the nineteenth century, Gaelic-speaking peasants were systematically driven from their land in order to give way to extensive sheep breeding. Today, Gaelic is a community language only in the Western Isles and on parts of Europe's western seaboard. Many linguistic phenomena encountered in Modern Irish, especially the Ulster Dialect, are found in Scottish Gaelic as well, including sound changes, mutations, and spelling.

Welsh

Written Old Welsh texts are attested from the ninth century. Until 1288, Welsh had status in administration and government. The Bible translations in 1567 (New Testament by William Salesbury and Richard Davies) and 1588 (the whole Bible by William Morgan) played a crucial role in Welsh history. Today, linguists distinguish two main dialects based on lexical variation: Northwalian Welsh and Southwalian Welsh, each comprising a number of more localized dialects, and the Literary or Formal Welsh language, itself comprising a number of written varieties or 'dialects'.

Welsh has a few distinctive consonant sounds, the most unusual being the voiceless lateral fricative *ll* and the voiceless alveolar trill *rh.* The unusual spellings *w* for [u] and *u* for [i] makes it seem as if there are too few vowels in a word. Three initial mutations, soft mutation, aspirate mutation, and nasal mutation, are present, and Welsh shares the VSO syntactic pattern: *Agorodd y dyn y drws* (lit. 'Opened the man the door').

Breton

Brittany was settled by people fleeing from Anglo-Saxons between the fourth and sixth century CE. Breton is therefore a Britannic language, even though Britanny is geographically European mainland. There is no unique standard written language. The main dialects are Cornouaille, Leon and Tregor, and Vannetais. Due to historical circumstances, Britanny has always been bilingual. As a consequence of permanent language contact, Breton has a very

French-sounding accentuation and sound system and also the strongest tendencies of all Celtic languages toward dropping its verb initial sentence structure.

Celtic Scholarship began in 1707 with the Welsh historian Edward Lhuyd (1660–1709) and his unfinished work *Archaeologia Britannica*. In 1853, Johann Kaspar Zeuss (1806–1856) published his *Grammatica Celtica*, which must be regarded as a further milestone in the establishment of Celtic Studies. A few more prominent Celticists were German native speakers: Ernst Windisch (1844–1918), Rudolf Thurneysen (1857–1940), Kuno Meyer (1858–1919), Julius Pokorny (1887–1970), and last but not the least Dane Holger Pedersen (1867–1953). Old Irish is most prominently studied, followed by Middle Welsh. Recent developments in European archeology have led to an increase in often-controversial works on Continental Celtic.

What all Celtic languages spoken nowadays have in common is that they are endangered minority languages, due to the predominance of English or, in Britany, French. Celtic linguistic history must be examined within the context of industrialization, and its negative influence on indigenous languages. Oral tradition is dying out in the wake of modern media, and today nearly all adult Celtic native speakers are bilingual. The British Labour Party has been regarded by some as the 'rescuer' of the Celts in Great Britain since it has engaged in decentralization and the founding of regional parliaments and assemblies. However, the future of the Celts as well as the languages remains to be seen.

References

Ball, Martin J. (ed.) 1993. The Celtic languages. London and New York: Routledge.

Black, Ronald, William Gillies, and Roibeard Maolalaigh (eds.) 1999. Celtic connections. Proceedings of the Tenth International Congress of Celtic Studies (Vol. 1, Language, literature, history, culture). East Linton: Tuckwell.

Durkacz, Victor Edward. 1983. The decline of the Celtic languages. Edinburgh: John Donald.

Hindley, Reg. 1990. The death of the Irish language. London and New York: Routledge.

Macaulay, Donald. 1992. The Celtic languages. Cambridge: Cambridge University Press.

Maier, Bernhard. 2000. Die Kelten. München: Beck.

McCone, Kim. 1992. Relative Chronologie: Keltisch. Rekonstruktion und relative Chronologie, ed. by Robert Beekes, Alexander Lubotsky and Jos Weitenberg, Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.

McCone, Kim. 1996. Prehistoric, old and middle Irish. Progress in medieval Irish studies, ed. by Kim McCone and Katherine Simms. Maynooth: Department of Old Irish.

Schrijver, Peter. 1995. Studies in British Celtic historical phonology. Amsterdam, Atlanta: Rodopi.

Sims-Williams, Patrick. 1998. The Celtic languages. The Indo-European languages, ed. by Anna Giacalone Ramat and Paolo Ramat. London and New York: Routledge.

Uhlich, Jürgen. 1999. Zur sprachlichen Einordnung des Lepontischen. Akten des zweiten deutschen Keltologen-Symposiums, ed. by Stefan Zimmer, Rolf Ködderitzsch and Arndt Wigger. Tübingen: Niemeyer

Untermann, Jürgen, and Dagmar Wodtko. 1997. Monumenta Linguarum Hispanicarum, Vol. IV (Die tartessischen, keltiberischen und lusitanischen Inschriften). Wiesbaden: Reichert.

ELVIRA M. VESELINOVIC

See also **Indo-European 1: Overview**

Chafe, Wallace

From the very beginning of his linguistic career Wallace Chafe has been involved in studies of American Indian languages, in particular, of the Iroquoian and Caddoan language families. In 1962, he published a general survey of the sociolinguistic state of the native languages of North America, indicating the number of speakers and the age of the youngest speakers, which remains relevant even now, 40 years later.

Among American Indian languages, Chafe has been involved most of all with Seneca, an Iroquoian language spoken in New York state, a polysynthetic language with the agent–patient system of role marking on the verb. Seneca examples cited in Chafe's work often demonstrate vast differences in conceptualization possible across human languages, especially as compared to European languages. Chafe also studied languages of the Caddoan language family, distantly related to Iroquoian.

In 1970, Chafe published a monograph entitled *A semantically based sketch of Onondaga* (Iroquoian) that was revolutionary for its time, being markedly different from both traditional American descriptivism with its emphasis on formal patterns and the growing

generative approach based on axioms of universal syntax. Chafe chose meaning as his starting point and presented the linguistic material as a system of mappings 'meaning => form'. In general, Chafe is unique among the American linguists of his generation in never being tempted to accept the generative Chomskyan approach.

This orientation toward semantics was also the chief element of his well-known book *Meaning and the structure of language* (1970), later translated into five languages. One of the main components of that book is a semantic classification of verbs. Along with Charles Fillmore, Chafe was among the first who realized the importance of semantic roles. The book also clearly demonstrated a typical feature of Chafe's subsequent publications — a very clear and simple style of writing that is of great persuasive force.

In the 1970s, the main focus of Chafe's theoretical research interests shifted to natural discourse. At that time, the trend that later acquired the label 'functionalism' and became a radical alternative to generativism started to take shape in the United States. Although Chafe was never an ideolog, he became one of the leading functionalists.

From the very beginning of his discourse studies, Chafe relied on a cognitive approach and tried to reconstruct the mental processes of a speaking person. In that respect, he was far ahead of the other students of text. In a widely known 1976 article, Chafe approached such familiar phenomena as referential devices, word order, and choice of subject from a cognitive perspective. His studies of the 1970s culminated in the 1980 monograph *The pear stories* in which an unusual method of data collection was used: Chafe and his younger associates (D. Tannen, J. du Bois, P. Clancy, among others) showed a short film of a boy collecting pears to speakers of different languages and ages who later retold what they saw. Thus, the input was controlled, and the different versions of the pear stories were available for comparison. Many innovative conclusions concerning processes of the verbalization of visual experience, the dynamics of human consciousness, cultural differences in emphasizing different layers of information, and the cognitive basis of grammatical choices were drawn in that study.

The following stage of Chafe's work is reflected in his recent (1994) book *Discourse, consciousness, and time.* This work looks into relationships between language and consciousness, the latter understood by Chafe as the mental system crucially responsible for discourse production. According to Chafe, language and consciousness cannot be understood and explored in isolation from each other. Chafe evaluates both disciplines studying those two domains, that is, modern linguistics and psychology, in a very critical vein, in particular, for their reliance upon artificial data and distrust toward introspective methods of research. Chafe emphasizes the use of introspection as a source of insights into language and cognitive processes, as well as the priority of natural data. The latter concern results from Chafe's inherent connection with the American ethnographic tradition that has always emphasized the importance of natural empirical data.

In this 1994 book, Chafe explores two main problems. The first is the explanation of linguistic phenomena on the basis of processes occurring in the speaker's consciousness. Consciousness can be immediate (a reflection of what is here and now) and displaced (memories, imagination). Immediate consciousness is more basic and elementary, while displaced consciousness can only be understood as a modification or complication of immediate consciousness. The second problem is the comparison of spoken and written language. Spoken language is more basic and universal, despite the fact that for decades linguists have directed their attention to written or quasi-written language only. Chafe gives priority to the spoken mode, but has also studied properties that distinguish written language.

The fundamental unit of discourse, according to Chafe, is the intonation unit, which is a quantum of discourse corresponding to one focus of consciousness. Each intonation unit typically contains one element of new information. The opposition 'given vs. accessible vs. new' is responsible for prosodic (stressed vs. unstressed) and lexical (pronoun vs. noun) realizations of referents. The status of syntactic subject, central for English grammar, is explained on the basis of the notion of the starting point for the expression of a focus of consciousness. The important light subject constraint suggests that subjects are normally given or accessible, and can be new only under very special conditions. The intermediate status of accessible information is the base for defining a discourse topic, which, in its turn, underlies a definition of sentence.

Among various modifications related to written language and the displaced mode of consciousness, prose written in the third person and quoted speech are discussed.

At present, Chafe is engaged in producing descriptions of the Seneca and Caddo languages, and he is also investigating the uses of prosody in ordinary speech, particularly for the expression of emotions and humor, and relations between language and music.

Among the main characteristics of Chafe's style of scientific thinking are the ability to look afresh at ordinary phenomena and to integrate a broad range of facts into a coherent picture, as well as independence from the dominant trends of thought. There is no doubt that Chafe is among the leading figures and a classic of modern linguistics, but like Sapir he has not created a circle of immediate followers and his works are not in the mainstream of modern linguistics.

Biography

Wallace Chafe was born in Cambridge, Massachusetts on September 3, 1927. He served in the US Navy, 1945–1946, and Central Intelligence Agency, 1950–1954. He received his B.A. in German from Yale University in 1950 and M.A. in linguistics from Yale University in 1956. He received his Ph.D. in linguistics from Yale University in 1958 for a dissertation on the Seneca language. He was Assistant Professor of Modern Languages at the University of Buffalo, 1958–1959, and linguist in the Bureau of American Ethnology, Smithsonian Institution, 1959–1962. In 1962, he moved to the University of California at Berkeley and worked there in the Department of Linguistics until 1986. From 1986 onward, he worked in the Department of Linguistics at the University of California at Santa Barbara. At present, he is Professor Emeritus at UCSB. He was Fellow of the Center for Advanced Study in the Behavioral Sciences, Palo, 1976–1977. He held a number of additional teaching appointments, including the Linguistic Institute of the Linguistic Society of America, in the summer of 1995, Albuquerque, New Mexico. Chafe was editor of the Mouton Grammar Library, 1984–1991; Topic Editor, Oxford International Encyclopedia of Linguistics; President of the Society for the Study of the Indigenous Languages of the Americas, 1985–1986; Associate Director of the Linguistic Institute of the Linguistic Society of America, Georgetown University, 1985; co-Director of the Center for the Study of Writing, Berkeley, 1985–1986; and member of the Linguistic Society of America Executive Committee, 1987–1989. He was also Member of the advisory and editorial boards of: International Journal of American Linguistics; Anthropological Linguistics; John Benjamins Typological Studies in Language Series; International Pragmatics Association; Humor: International Journal of Humor Research; Text: An Interdisciplinary Journal for the Study of Discourse; and Consciousness and Cognition. In addition, he was a member of the following professional societies: Linguistic Society of America; American Anthropological Association; American Psychological Association; American Psychological Society; Society for Linguistic Anthropology; Society for the Study of the Indigenous Languages of the Americas; International Society for Historical Linguistics; and International Pragmatics Association. He was Skomp Distinguished Lecturer, Indiana University, 1997, and received the medal of the University of Helsinki in 1998.

References

Chafe, Wallace. 1962. Estimates regarding the present speakers of North American Indian languages. International Journal of American Linguistics 28.

———. 1967. Seneca morphology and dictionary. Washington: Smithsonian Press.

———. 1970. A semantically based sketch of Onondaga. International Journal of American Linguistics, Memoir 25 (Supplement to 36 (2)).

———. 1970. Meaning and the structure of language. Chicago: University of Chicago Press.

———. 1974. Language and consciousness. Language 50.

———. 1976. The Caddoan, Iroquoian, and Siouan languages. The Hague: Mouton.

———. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. Subject and topic, ed. by C.N. Li. New York: Academic Press.

——— (ed.) 1980. The pear stories: cognitive, cultural, and linguistic aspects of narrative production. Norwood, NJ: Ablex.

Chafe, Wallace, and Johanna Nichols (eds.) 1986. Evidentiality: the linguistic encoding of epistemology. Norwood, NJ: Ablex.

Chafe, Wallace. 1994. Discourse, consciousness, and time: the flow and displacement of conscious experience in speaking and writing. Chicago: University of Chicago Press.

———. 1996. Sketch of Seneca, an Iroquoian language. Handbook of North American Indians. Vol. 17, Languages, ed. by W.C. Sturtevant. Washington, DC: Smithsonian Institution.

———. 1998. How a historical linguist and a native speaker understand a complex morphology. Historical linguistics 1997, ed. by M.S. Schmid, J.R. Austin, and D. Stein. Amsterdam/Philadelphia: John Benjamins.

———. 1998. Language and the flow of thought. The new psychology of language: cognitive and functional approaches to language structure, ed. by M. Tomasello. Mahwah, NJ: Lawrence Erlbaum.

———. 2000. Loci of diversity and convergence in thought and language. Explorations in linguistic relativity, ed. by M. Pütz, and M. Verspoor. Amsterdam/Philadelphia: John Benjamins.

Chafe, Wallace. Autobiography. Historiographia Linguistica (in press).

Parret, Herman (ed.) 1974. Interview with Wallace Chafe. Discussing language. The Hague: Mouton.

Kibrik, Andrej A. 1994. Cognitive studies in discourse. Voprosy jazykoznanija 5 [in Russian].

ANDREJ KIBRIK

Chao Yuen Ren

A Chinese–American linguist and one of the leading figures in American Structuralism, Chao Yuen Ren made important contributions in the general theory of language and by introducing the Chinese language(s) as an object of research to a broader linguistic community. His work covers a wide range of issues both in theoretical and in applied linguistics, including fieldwork on Chinese dialects (usually conceived today as separate languages), the grammar of Modern Chinese, Chinese lexicography, practical material for Chinese language teaching, language standardization in China, the general theory of language, and studies in sociolinguistics.

Born into a family belonging to the old scholar and mandarin class, Chao's interest in linguistic variability was aroused during early childhood, when he came into contact with, and learned to speak, a number of different dialects. His family originally came from the town of Changzhou (between Nanjing and Shanghai), where the westernmost variety of the Wu dialect group is spoken. During the first half of his childhood, however, the family lived in various places in Hebei province, Northern China, and frequently moved around between the then provincial capital Baoding and smaller places. In those years, Chao was therefore exposed not only to the Wu dialects spoken by the older generations of his family but also to different varieties of Northern Mandarin, which he heard from his surroundings.

Although he majored in mathematics and physics at Cornell University, he increasingly took an interest in philosophy and was influenced in particular by the writings of Bertrand Russell. However, he kept up his strong interest in languages, further stimulated by a course in phonetics and other linguistic courses. During those years, he also learned German and French, and picked up more Chinese dialects from Chinese fellow students. During his graduate studies in the Department of Philosophy at Harvard, he attended an introductory course in linguistics and took quite a few language courses on the side, among them a course in Sanskrit.

During the winter of 1920/1921, Chao served as an interpreter to Bertrand Russell during his lecture tour in China and made full use of his mastery of the dialects. This experience provided a challenge to Chao in a way physics never had, and it was then that he decided upon linguistics as a major field of study and research. Therefore, while teaching philosophy at Harvard during the next three years, he took linguistic courses at the same time, and subsequently went to Europe to work and study with the phoneticians Daniel Jones and Stephen Jones, the linguists Joseph Vendryes and Antoine Meillet, and the sinologists Bernhard Karlgren, Paul Pelliot, and Henri Maspero.

After his return to China, primarily to teach Chinese phonology, he spent more than ten years making and organizing surveys of the Chinese dialects, which resulted in a number of pioneering studies: the first one was on the Wu dialects of Jiangsu and Zhejiang provinces (1926–1927), published as his ground-breaking *Studies in the modern Wu dialects* (1928), followed by surveys on the Yue dialects of Guangdong and Guangxi provinces (1928–1929), Minnan dialects of Shantou and Chaozhou and the dialects of Anhui (1934), Jiangxi and Hunan (1935), and Hubei provinces (1936). During the late 1920s, he was also active in the Committee on Unification of the National Language and headed a group of Chinese linguists who developed a new system of Latinization for Chinese, National Language Romanization (*Gwoyeu Romatzyh*), which was unique in using spelling variants instead of diacritics to express the tones of Chinese.

Chao had plans for a dialectal survey of the whole country, but his fieldwork was eventually interrupted by the political development in China and the outbreak of war with Japan. Following these events, in 1938 he came to the United States for good and subsequently taught at the University of Hawaii; at Yale, where he met Leonard Bloomfield, Edward Sapir, and Bernard Bloch; at Harvard; and, finally, at the University of California, from where he retired in 1960.

Chao's pioneering fieldwork on the Chinese dialects is clearly one of his main contributions to linguistics. Before him, the European sinologists Henri Maspero (1883–1945) and Bernard Karlgren (1889–1978) had used Chinese dialect data in comparative studies for reconstructing earlier stages of Chinese (in particular, what Karlgren called 'Ancient Chinese'), but they were interested in Chinese dialects only to the degree that they had preserved the phonetic features of those earlier stages. Chao, however, was the first to look at the Chinese dialects from a truly modern perspective, recognizing them as objects worth studying for their own sake. He carried out wide-ranging dialect surveys using the most modern sound recording devices of the times, and he included real dialect words found in nonlearned, colloquial speech. Although Chao's fieldwork was disrupted by the historical events in China and thus remained

fragmentary, it provided a solid foundation for later work, much of which was done by his students.

In the field of Chinese grammar and language teaching, Chao wrote two epoch-making books in which he applied the grammatical approach of American structuralism to the description of Modern Chinese: *Mandarin primer,* published in 1948 as a textbook for American students learning Chinese, contains an introductory grammatical sketch that was translated into Chinese and had a tremendous impact in China, effectively introducing the methodology of structuralism to a wide Chinese audience. The grammatical sketch of the *Primer* provided the best linguistic description of Modern Chinese until 1968, when Chao published *A grammar of spoken Chinese*, a monumental work that was the first comprehensive reference grammar of Chinese written in English, and that is arguably the best description of Modern Chinese to this day.

In Chinese lexicography, Chao and Yang's *Concise dictionary of spoken Chinese* (1947) made a number of important lexicographic innovations, restricting entries to those morphemes that occur in the modern spoken or written language, marking the morphemes as 'free' or 'bound' from the point of view of the modern language, and as 'literary' if they occur in the written language only. Pronunciation of the entries is given in National Language Romanization.

Among Chao's works on theoretical phonology, the most influential one was his classical paper 'The non-uniqueness of phonemic solutions of phonetic systems' (1934/1957), which deals with the definition of the phoneme and tries to show that the principles for establishing a system of phonemes for a given language depend on a number of possible objectives set up by the linguist, some of which are external to the language system itself. Chao also introduced a convenient device of representing tones in phonetic or phonological transcriptions, which became known as 'tone letters' and was later adopted for the International Phonetic Alphabet. Over the years, a number of important articles on the problems of phonetics and transcription followed.

More than anyone else, Chao laid the foundation for studying Chinese within the international perspective of modern linguistics, and played a key role in introducing the ideas of Western linguistics to linguistic scholars inside China. Outside the area of linguistics, Chao is remembered as an outstanding translator (*Alice's adventures in wonderland,* 1922) and as a composer of Chinese and Western music.

Biography

Chao was born in Tianjin, China, on November 3, 1892. From the age of seven, he was given a traditional Chinese education in the family school at home. The family returned to Changzhou in 1901, and after both his parents died in 1904, he spent a year in his aunt's home in Suzhou. After attending a primary school in Changzhou in 1906, he was sent to Nanjing for three years to continue his studies. In 1910, he passed a government examination, which earned him a scholarship for studying in the United States of America, and he went to the Tsing Hua College in Beijing to prepare himself. Chao thus studied mathematics and physics at Cornell University, Ithaca, B.A. (1914), followed by physics and philosophy at Harvard, where he earned his Ph.D. degree in 1918 with *Continuity, a study in methodology*, which was supervised by the philosopher and logician Henry Maurice Sheffer. Chao received the post-doctorate Sheldon Fellowship, 1918–1919 and went on to teach physics at Cornell in 1919–1920. He then went back to China to teach mathematics at Tsing Hua College in Beijing, but was soon asked to interpret for Bertrand Russell on his lecture tour of China, 1920–1921. Afterwards, he returned to Harvard to teach philosophy and Chinese, at the same time taking linguistic courses (1921–1924). He went to Europe to visit the sinologist Bernhard Karlgren in Gothenburg, took courses in phonetics with Daniel Jones and Stephen Jones in London, and attended lectures by Joseph Vendryes, Antoine Meillet, Henri Maspero, and Paul Pelliot in Paris (1924). Returning to China, he taught Chinese phonology and music at Tsing Hua College's newly established Institute of Sinology (1925), and then conducted and organized fieldwork on Chinese dialects between 1926 and 1936. He was an active member of the Committee on Unification of the National Language and the chairman of the Linguistics Section of the Institute of History and Philology, Academia Sinica (1929–1938). He was visiting professor at the Sunyatsen University, Canton (1928—1929) and then temporarily returned to the United States of America as the director of the Chinese Educational Mission in 1932–1933. He obtained his American citizenship in 1938 and then taught literary Chinese as a visiting professor at the University of Hawaii (1938–1939) and Yale (1939–1941) before returning to Harvard once again, as a researcher with the Harvard-Yenching Institute's Chinese–English Dictionary Project. He also taught Chinese and Cantonese, and served as head of Harvard's wartime Chinese language program (1941–1947). He became Professor of Oriental Languages and Linguistics at the University of California, Berkeley in 1947 and was made Agassiz Professor of Oriental Languages and Linguistics in 1952. A guest lecturer at National Taiwan University and Kyoto University, Japan in 1959, he retired the following year, in 1960.

Chao was also President of the Linguistic Society of America, 1945; full member of the Academia Sinica, 1947; Fellow of the American Academy of Arts

and Sciences, 1948; and President of the American Oriental Society, 1960. He earned honorary degrees from Princeton University (1946), the University of California (1962), and Ohio State University (1970) and was named Honorary Professor at Beijing University in 1981. Chao died in Cambridge, Massachusetts, on February 24, 1982.

References

Bibliography of Yuen Ren Chao's works, In Chao. 1976. 402–15.
Chao Yuen Ren. 1928. Xiandai Wuyu de Yanjiu (Studies in the modern Wu dialects).
———.1930. A system of tone letters. Le Maître Phonétique 45. 24–7.
———.1934. The non-uniqueness of phonemic solutions of phonetic systems. Bulletin of the Institute of History and Philology (Academia Sinica) 4. 363–97; reprinted in Readings in linguistics, ed. by Martin Joos, 38–54. Washington, DC: American Council of Learned Societies, 1957.
———.1947. Cantonese primer.
———.1948. Mandarin primer.
———.1968. A grammar of spoken Chinese.
———.1968. Language and symbolic systems.
———.1968. Readings in sayable Chinese.
———. 1975. Yuen Ren Chao's autobiography: first 30 years, from 1892–1921. Ithaca, New York: Spoken Language Services.
———. 1976. My linguistic autobiography. In Chao. 1–20.
———. 1976. My fieldwork on the Chinese dialects. In Chao. 26–33.
———. 1976. Aspects of Chinese sociolinguistics, ed. by Anwar S. Dil. Stanford: Stanford University Press.
Chao, Yuen Ren. and Yang Lien-Sheng. 1947. Concise dictionary of spoken chinese.
Dil, Anwar S. 1976. Introduction. In Chao. xi–xiv.
Egerod, Søren, 1996. Chao Yuen Ren. Lexicon Grammaticorum, ed. by Harro Stammerjohann et al., 178–9. Tübingen: Niemeyer.
Wang, W. S.-Y. 1983. Chao Yuen-Ren. Language 59. 605–7.

ROBIN SACKMANN

See also **Structuralism**

China

The region of China discussed in this essay covers the entire territory claimed by the People's Republic of China, including Taiwan. It consists of 23 provinces, five autonomous regions, four municipalities, and two special administrative regions in an area of over 9,630,000 square kilometers. With a population of 1,295,330,000 (in the 2000 census), China is the most populous country and one of the most diversified regions in terms of languages and cultures in the world. No accurate number for languages spoken in China is available, but at least 235 languages (including signed languages and ten extinct languages) have existed in this region, cf. Grimes's *Ethnologue* (2000). Barring a few exceptions, they belong to one of the following groups (parenthetical figures indicate number of members and the group's percentage of total living languages): Sinitic (13, 5.78%), Tibeto-Burman (72, 32%), Tai-Kadai (31, 13.78%), Hmong-Mien (32, 14.22%), Altaic (25, 11.11%), Austro-Asiatic (24, 10.67%), Austronesian (18, 8%), and Indo-European (4, 1.78%). Figure 1 outlines their approximate distribution in China.

The map shows that Sinitic sits in the center, but it has spread out to every corner of China. There has been a general migration flow of people from the north to the south throughout much of the history of China. Southward invasion by Altaic speakers sometimes leads to the spreading of Chinese further north, as in the case of Manchu's government. On the other hand, southward expansions and escapes by Sinitic speakers are responsible, in part, for the migration of Hmong-Mien and Tai-Kadai speakers to their present-day habitation.

Scholars in China generally consider the Sino-Tibetan family to consist of four branches: Sinitic, Tibeto-Burman, Tai-Kadai, and Hmong-Mien (Miao-Yao). While the typological profiles of the latter two are rather similar to that of Sinitic, their genetic relationship with the family remains to be proven. Table 1 summarizes three aspects of the major language groups as to whether their members (confined to those in China) are tone languages, whether morphemes (the smallest unit that carries a meaning) in the languages are predominantly monosyllabic, and in what order the Subject (S), Verb (V), and Object (O) appear in canonical sentences.

The boundary between language and dialect is fuzzy in nature, and highly controversial when political factors are taken into consideration. The case of Chinese is a good example of this. The Chinese identify themselves ethnically as *Han*, and *Hanyu* is the
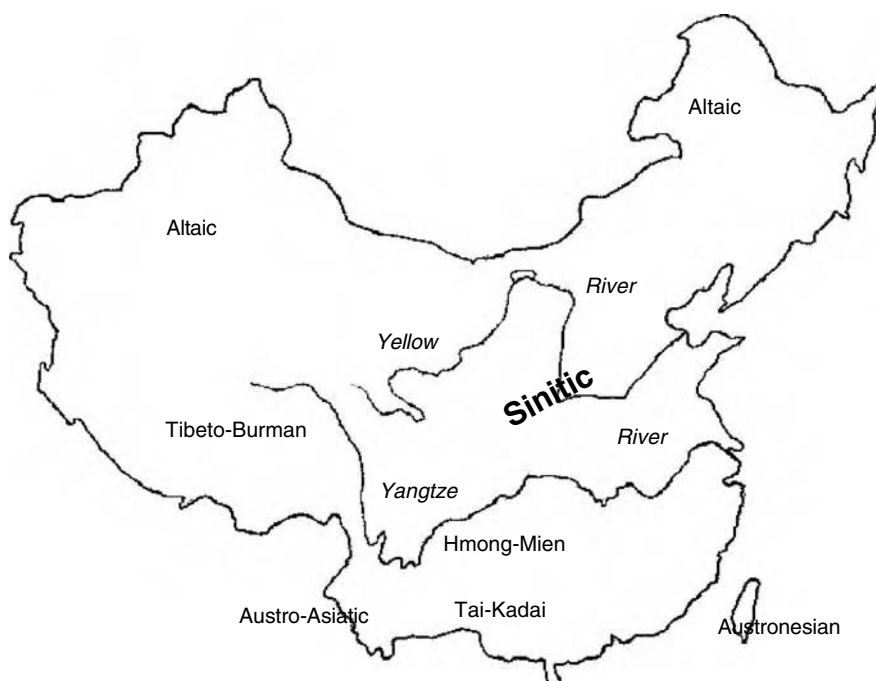
Figure 1. The distribution of major language groups in China.

TABLE 1 A Brief Typological Comparison of the Major Language Groups in China

Language Group	Tone Language	Monosyllabicity of Morphemes	Word Order
Sinitic	All	Yes	S–V–O
Tibeto-Burman	Most	Yes	S–O–V
Tai-Kadai	All	Yes	S–V–O
Hmong-Mien	All	Yes	S–V–O
Altaic	None	No	S–O–V
Austro-Asiatic	Few	Most	S–V–O
Austronesian	None	No	V–S–O

official language of China. From a strictly linguistic point of view (based on mutual intelligibility), however, at least 13 Chinese languages are spoken in modern China (see Figure 2). In the Chinese tradition, these are construed as dialects of *Hanyu* or subdialects therein. Prior to the twentieth century, Classical Chinese had been used in writing for all purposes, serving as the official language of the emperor for many dynasties. At the May 4th movement in 1919, student protesters in Beijing held among their goals a standard written language based on the modern Chinese speech. The then Government of the Republic of China responded by designating *Gwoyu* 'national language', based on Beijing Mandarin, as

the standard language of the new republic. Later in the 1950s, the Government of the People's Republic of China promoted *Putonghua* 'common language' (also based on Beijing Mandarin) as the national language. *Gwoyu* (still in use in Taiwan) and *Putonghua* have little difference in phonology. Their divergence has largely resulted from further language standardization introduced on the mainland, notably the replacement of phonetic characters by *pinyin* in the Roman alphabet and especially the use of simplified characters. The latter has compromised the antediluvian tradition of a unified written form for the Chinese since the unification of the warring states in China in 221 BCE.

Encyclopedia Sinica (1988) estimates the number of speakers of seven major dialect groups of *Hanyu* in the early 1980s as follows: over 700 million for Mandarin, 70 million for Wu, 40 million for Yue, 40 million for Min, 25 million for Xiang, 30 million for Gan, and 37 million for Hakka. Figure 2 shows the relation between 13 identified Sinitic languages. The tentative placement of Wan Nan is geographically motivated.

The Yangtze River is the watershed between Northern Chinese and Southern Chinese. Mandarin stretches from the vast region north of the Yangtze River to southwestern China provinces of Sichuan, Yunnan, Guizhou, western Hunan, and northwestern Guangxi. Following the escape of the Nationalist government from the mainland, a sizable number of Mandarin speakers settled down in Taiwan in the
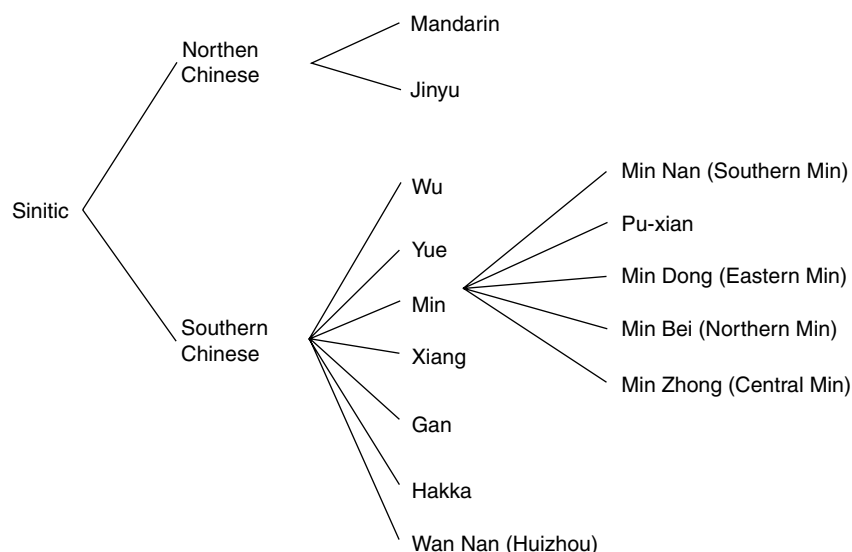
Figure 2. A linguistic tree of identified Sinitic languages in China.

1950s. More than 870 million people (in the late 1990s) speak Mandarin as a first language, including nearly all Manchu and the entire population of Hui, the second and third largest minority nationalities of China. The figure rises well beyond one billion if the number of fluent speakers of Mandarin as a second language is taken into account. Mandarin has four major dialects: Northern Mandarin, Northwestern Mandarin, Southwestern Mandarin, and Lower Yangtze Mandarin.

Found in coastal provinces, Wu is spoken on the east coast, Yue on the south coast, and Min on the southeast. The Yangtze Delta can be considered as the homeland of Wu. Wu is spoken nowadays in Zhejian province and much of the area south of the Yangtze River in Jiangsu province. It has five dialects: Taihu (which includes Shanghai as one of its six subdialects), Taizhou, Wenzhou, Wuzhou, and Liqu. The homeland of Yue (Cantonese) is in the Pearl Delta; from there, it spreads to a large part of Guangdong province and as far as southeastern Guangxi. Yue dialects include Yuehai (which is spoken in Guangzhou, Hong Kong, and Macau), Siyi, Gaoyang, and Guinan.

Min is regarded as the fourth largest dialect of *Hanyu*, and also the most diversified one with respect to dialectal difference. Indeed, the diversion is so significant that it warrants the treatment of Min as a linguistic group comprising five consanguineous, but distinct, languages. Li's *Dialects of Fujian* (1997) describes the distribution of these languages as follows: excluding a small Wu-spoken area in the northern corner and about one third of the province in western Fujian (where Hakka and Gan are spoken), the rest of Fujian can be divided into six areas: the south-

ern half contains Min Nan and Pu-xian (the latter occupies only a corner in the east), and the northern half contains Min Dong (in the eastern half), Min Bei (about 60% of the remainder in the north), and Min Zhong (the western half of the leftover). A transitional language between Min Nan and Min Dong exists in Youxi county, the very center of the province.

Min Nan is the largest Min language, with 25.7 million speakers (in 1984) in southern Fujian, eastern and western parts of Guangdong, parts of Hainan and Zhejiang, and an additional 15 million speakers (in 1993) in Taiwan. Major Min Nan dialects include Hokkien, Chao-Shan, Leizhou, Hainan, and Zhenan Min. Many can be subdivided, e.g. Hokkien has four subdialects: Amoy, Quanzhou, Zhangzhou, and Longyan. Pu-xian and Min Zhong occupy the smallest areas among the Sinitic, but Pu-xian has around three million speakers, whereas speakers of Min Zhong are less than one million. They each have two dialects, and so do Min Dong and Min Bei.

Xiang is mainly spoken in much of the eastern half of Hunan province. With Mandarin dominant in the north and in the west, Xiang has been under great pressure of assimilation. The influence of Mandarin represents a primary factor for variation between its two dialects: Northern Xiang is subject to more innovation, while Southern Xiang is more conservative. Gan is largely spoken in Jiangxi province, with five major dialects: Chang-Jing, Yi-Liu, Ji-Cha, Fu-Guang, and Ying-Yi. It is the closest relative of Hakka.

The distribution of Hakka is unique among Southern Chinese in that it covers the widest area in China, but often with a low concentration; it does not have a home base in any province. While the border area between northeastern Guangdong, southern

Jiangxi, and southwestern Fujian has a higher concentration of Hakka speakers, small Hakka-speaking communities are scattered in parts of Taiwan, Hunan, Sichuan, Guangxi, and throughout Guangdong. Major Hakka dialects include Min-ke, Yue-Tai, Yuezhong, Yuebei, Yugui, Tonggu, etc. The majority of She nationality also speak Hakka as their first language.

Jinyu and Wan Nan are two of the less-known Chinese languages. Spoken in Shanxi province and adjacent areas in Shaanxi and Henan provinces, Jinyu is generally assumed to be a subdialect of Northwestern Mandarin. Wan Nan is mainly spoken in an area called Huizhou in southern Anhui province. It is usually treated as a subdialect of Lower Yangtze Mandarin.

The Chinese government recognizes only 55 minority nationalities. Their total population is about 106.8 million, or 8.25% of China's population (in the 2000 census). Around 60% of them are native speakers of over 94% of languages in China. Table 2 shows the top 12 minority languages, each with one million or more speakers (in the early 1980s or 1990s).

Spreading across Xizang Tibetan Autonomous Region, southeastern Qinghai, western Sichuan, and much of the Yunnan provinces, Tibeto-Burman accounts for almost one third of identified languages in China. The three largest groups within it are Yi-Burmese (28 languages), Himalayish (17 languages), and Qiangic (15 languages). Speakers of Yi-Burmese in China exceed 6.4 million, including such nationalities as Yi, Hani, Lisu, Lahu, Naxi, Nu, Jino, and Achang. Chinese scholars consider that Yi nationalities speak six rather distinct "dialects" (just as the Chinese themselves do). Yi nationalities use an ideographic system for writing their languages, but characters vary considerably from district to district over southeastern Sichuan, Guizhou, and Yunnan. A set of

standard characters based on Sichuan Yi was introduced in 1975. Tomba is the traditional writing developed by Naxi nationality. The hieroglyphic writing, however, is restricted to religious texts only.

Himalayish languages have about 4.1 million speakers in China. Over 82% of them speak Tibetan, Khams, or Amdo. These three consanguineous languages (regarded as 'dialects' of Tibetan in China) share a common writing script adapted from the Indic syllabic writing system. Besides Tibetan nationality, the other official nationality for speakers of Himalayish is Moinba (whose population was less than 8,000 in the 1990 census). Exclusive to China, Qiangic languages have about 490,000 speakers in total. Approximately one fifth of Qiangic speakers are Qiang or Pumi nationalities; the rest identify themselves officially as being of Tibetan nationality. Historically, their languages have no written form, with the exception of Tangut (an extinct relative), which had a logographic writing modeled after Chinese characters. Bai, Tujia, Jingpo, Drung, and Lhoba are among the 17 official nationalities speaking Tibeto-Burman languages in China.

Spoken in much of Guangxi Zhuang Autonomous Region, Hainan, southern Guizhou, and southwestern Yunnan, Tai-Kadai has the second greatest number of speakers in China. Chinese scholars regard Northern Zhuang and Southern Zhuang as dialects of Zhuang, which was once written in logographic characters akin to Chinese characters, but a new Roman orthography was introduced in the late 1950s. Together with 20 others, the Zhuang languages belong to the Kam-Tai group. Seven nationalities are designated for the 20 million speakers of Kam-Tai, namely Zhuang, Buyi, Dai, Dong, Sui, Mulam, and Maonan. The Dai have adapted Brahmi script (a syllabic Indic script) for writing Lü and Tai Nüa for many centuries. Kadai (81,000

TABLE 2 Top Twelve Minority Languages of China

Language Name	Linguistic Affiliation	Nationality	No. of Speakers
Northern Zhuang	Tai-Kadai, Kam-Tai	Zhuang	10,000,000
Uyghur	Altaic, Turkic	Uygur	7,200,000
Southern Zhuang	Tai-Kadai, Kam-Tai	Zhuang	4,000,000
Mongolian	Altaic, Mongolic	Mongol	3,381,000
Bouyei	Tai-Kadai, Kam-Tai	Buyi	2,000,000
Korean	Altaic	Korean	1,920,000
Sichuan Yi	Tibeto-Burman, Yi-Burmese	Yi	1,600,000
Khams	Tibeto-Burman, Himalayish	Tibetan	1,487,000
Southern Dong	Tai-Kadai, Kam-Tai	Dong	1,480,750
Kazakh	Altaic, Turkic	Kazak	1,100,000
Tibetan	Tibeto-Burman, Himalayish	Tibetan	1,066,200
Hmong Njua	Hmong-Mien, Hmongic	Miao	1,000,000

speakers in five languages) and Hlai (747,000 speakers in two languages) are two smaller groups within Tai-Kadai. Gelao and Li are the two official nationalities speaking Kadai and Hlai, respectively.

All 32 Hmong-Mien languages found in the world have speakers in China, scattered in various parts of Guizhou, Hunan, Guangdong, Guangxi, and Yunnan. The three official nationalities—Miao, Yao, and She—correspond to the three language groups within Hmong-Mien. Hmongic is the largest, comprising 26 languages spoken by 4.2 million people who are mostly Miao, but some are Yao. Over 1.1 million Yao speak five Mienic languages, but less than 1,000 She speak the singleton language in the Ho Nte group. Hmong-Mien languages have no traditional writing, but Roman orthographies have been devised for several larger languages since the 1950s.

With 11 Turkic languages, eight Mongolic languages, and five Tungusic languages, all three major branches of the Altaic family are present in China. Spoken along the northern and northwestern border of China—mostly in Heilongjiang province, Inner Mongolian Autonomous Region, and Xinjiang Uygur Autonomous Region—by 18 nationalities, the Altaic languages amount to 25, if Korean is included. Khakas, Nanai, and Manchu each have less than 50 speakers in China. According to estimates of the 1990s, Turkic has over 8.49 million speakers among Uygur, Kazak, Kirgiz, Salar, Bonan, Ozbek, Tatar, and Yugur nationalities; Mongolic has nearly 4.15 million speakers among Mongol, Tu, Daur, Dongxiang, and Yugur nationalities; and Tungusic has about 49,700 speakers among Xibe, Ewinki, Oroqen, Hezhen, and Manchu nationalities. Arabic script is used for writing Uyghur, Kazakh, and Kirghiz, whereas traditional Mongolian script is used for writing Mongolian and Xibe. Koreans write in Hangul, a phonetic-based spelling system created in 1444.

Va, Blang, and De'ang nationalities in western Yunnan, plus Jing nationality on the south coast of Guangxi, are speakers of Austro-Asiatic languages. All 24 Austro-Asiatic languages of China belong to the Mon-Khmer branch, with around 381,000 speakers in total. Only one sixth of them have more than 10,000 speakers. Parauk is the largest, with about 180,000 speakers (in 1990).

The mainland government designates an Austronesian-speaking minority in Taiwan as Gaoshan nationality, but on the island they are further divided into nine ethnic groups, collectively referred to as Aborigines. The total number of Austronesian speakers in China is around 339,800 for 18 Austronesian languages. Those spoken on the Taiwan island belong to the Formosan branch, which comprises three groups: Paiwanic (ten languages), Tsouic (four languages), and Atayalic (two languages). The largest is

Amis (with 130,000 speakers) of the Paiwanic group, which lost seven members during the colonial rule of Japan (from 1895 to 1945). Grimes's *Ethnologue* reports that another four Paiwanic languages and two Tsouic languages are nearly extinct. Unlike mainland China, Taiwan has witnessed language extinction similar to that found in Australia and North America, in the violent way of supplanting an aboriginal language with a colonial one. Under the rule of the Nationalist government in the 1950s, linguistic imperialism was imposed on the Taiwanese people with Mandarin in lieu of Japanese. As the Taiwan government has advocated multiculturalism in recent years, the environment has began to change to one propitious for linguistic diversity. Yet, many educated Chinese parents in Taipei (the capital city) prefer speaking Mandarin to their young children, even though they are native speakers of Min Nan. Mandarin has undoubtedly established its superior status in Taiwan over the half century.

The firmly established status of Mandarin is also unchallengeable in mainland China. While the Chinese government does not adopt any aggressive language policies to wipe out linguistic minorities, policies that would sustain linguistic diversity are equally lacking. It is stated in the Chinese Constitution that minority nationalities have the right to maintain their own languages, but this by no means implies active promotion of minority languages. In spite of the availability of newly devised orthographies for languages of a few larger nationalities, the priority for mastering *Putonghua*, which is often the sole language in education and mass media, cannot be mistaken. Ethnic minorities have been in a plight parallel to that of new immigrants in countries such as the United States. In order to become part of the mainstream society, they have to give up their native languages, sometimes eagerly. Interethnic marriage (not necessarily with *Han*-Chinese) in cities is another factor for prompting children to become monolingual in Mandarin. This kind of language shift is usually completed within several generations, characterized by bilingual speakers during the peaceful transition.

Bilingual speakers of autochthonous languages are legion in China but the bias toward Mandarin is conspicuous. Most of the bilinguals, except for a small fraction of minority nationalities, are literate only in standard Chinese based on Mandarin. Linguistic minorities surrounded by larger communities of different languages are likely to become multilingual in several oral languages, e.g. some Zhuang nationalities speak Yue and Mandarin in addition to their native language and some Pumi nationalities are fluent in Prinmi, Mandarin, and Sichuan Yi. While considerable ethnic minorities (especially those above the age of 50) are fluent in two or more languages, the ratio of

monolingual Mandarin speakers in minority nationalities is steadily rising.

Ethnic minorities are not alone under the pressure of linguistic absorption. All languages other than Mandarin are facing varied degrees of endangerment, as *Putonghua* is upheld as the national language at the expense of the others. For instance, vernacular pronunciation of Chinese characters is no longer taught in school (except in Hong Kong and Macau). In the past, Wu, Yue, Min Nan, and other Sinitic languages were featured in vernacular operas in different parts of the country, but nowadays popular songs are sung exclusively in *Putonghua* in mainland China. These suggest an ever-shrinking room for the survival of other languages in China.

The linguistic landscape in China is byzantine. Many languages are awaiting in-depth studies and proper classification, while some are yet to be discovered. Hopefully, these would be accomplished before language death accelerates under the momentous impact of the on-going economic development and modernization.

References

Grimes, Barbara (ed., Joseph Grimes, consulting editor) 2000. Ethnologue: languages of the world, 14th edition. Dallas, TX: Summer Institute of Linguistics.

Ji, Xianlin (editor-in-charge) 1988. Zhongguo Da Baikequanshu: Yuyan, Wenzi (Encyclopedia sinica: languages and scripts). Beijing: China Encyclopedia Publisher.

Li, Rulong. 1997. Fujian Fangyan (Dialects of Fujian). Fuzhou: Fujian People's Publisher.

Norman, Jerry. 1998. Chinese. Cambridge: Cambridge University Press.

Ramsey, Robert. 1987. The languages of China. Princeton, N J: Princeton University Press.

Wang, William S.-Y. (ed.) 1991. Languages and dialects of China. Berkeley, CA: Journal of Chinese Linguistics.

Wurm, Stephen, et al. (general editors) 1987. Language atlas of China. Hong Kong: Longman.

Wurm, Stephen, Peter Muhlhausler, and Darrell Tryon (eds.) 1996. Atlas of languages of intercultural communication in the Pacific, Asia, and the Americas. Berlin and New York: Mouton de Gruyter.

PICUS SIZHI DING

See also **Altaic; Austronesian; Chinese (Mandarin); Chinese and Japanese Traditional Grammar**

Chinese (Mandarin)

The Chinese language is represented by groups of dialects that are approximately as distant from each other as English from German, or French and Spanish. The traditional classification has been into seven major groups: Mandarin, Wu, Xiang, Gan, Min, Yue, and Hakka. Collectively, these dialects are spoken by more speakers than any other language in the world. Of these, Mandarin has the special status of having been the official language of China for most of Chinese history. As such, Mandarin is the first dialect learned by the great majority of people in China, and is acquired to various degrees of proficiency by speakers of all other dialects. Thus, there are well over a billion people who use Mandarin in some way or other in their lives.

Whereas the other dialects are concentrated mostly along the southeastern coast, Mandarin is distributed widely all over China. The Mandarin spoken in Harbin in the northeast is clearly different from the Mandarin spoken in Chengdu in the southwest, a thousand miles away. Because Beijing has been the capital of China for most of the past millennium, its speech has become an official linguistic standard, recognizing of course that the term Mandarin in its wide sense actually covers many varieties of speech.

There are many ways to refer to this language. Traditionally, when there is no need to differentiate it from the other major dialects, Mandarin is simply called 'Chinese'. We will follow this tradition in the present article. The word 'Mandarin' actually means 'official', corresponding to the Chinese word 'guan'. Thus, the Chinese counterpart to 'Mandarin' as a language is 'Guanhua', or 'official speech'. To emphasize its wide usage and enhance its populist intent, the term currently preferred is 'Putonghua', where 'putong' means 'common'. The term 'Guoyu', or 'national language', is used in Taiwan. The term 'Huayu' is used in Singapore, where 'hua' is an auspicious word that has been used to designate ethnic Chinese since the ancient Xia dynasty, some 4,000 years ago. In fact, the two elements 'hua' and 'xia' are often used together as a compound to refer to the long ethnic heritage.

Lastly, the term 'Sinitic' is used in more technical discussions, such as 'the Sinitic language', meaning 'the Chinese language'. The stem 'Sin-' probably derives from the name of the Qin dynasty (221–206

BCE), when China was first unified politically. 'Sino-Tibetan' refers to the family of languages one branch of which is Sinitic; another branch is Tibeto-Burman, which contains several hundred languages spoken in Western China and around the Himalaya mountains. The traditional classification of Sino-Tibetan is that proposed by F.K. Li; it includes a third branch of Zhuang-Dong languages, and a fourth branch of Miao-Yao languages. This four-branch hypothesis has recently become controversial as more data on minority languages became available, although the relation between Chinese and Tibeto-Burman has never been disputed. It has been estimated that the Chinese and Tibet-Burman retained their unity until about 6,000 years ago, when extensive migrations radiated outward from its homeland, presumably in northern China, which ultimately gave rise to the hundreds of languages we see today.

There have been efforts to connect the Sino-Tibetan languages to much more distant origins. In the early part of the twentieth century, the American linguist Edward Sapir remarked on the deep similarities between Tibetan and some of the native languages of North America. More recently, research is being done on the Dene-Caucasian hypothesis, which posits that groups as distant as the Na-Dene languages of North America, the Yeniseian languages in Siberia, Basque in Western Europe, among others, are all related to Sino-Tibetan.

A competing hypothesis links Chinese to the Austronesian languages, thus implying an ancient homeland more to the south. Such hypotheses on distant prehistories are intrinsically fascinating, although naturally they are more difficult to verify, often for lack of adequate data or precise methodology. An encouraging development in recent years is the increase in interdisciplinary collaboration, particularly among linguists, anthropologists, and geneticists, which significantly broadens our perspectives on human evolution, and brings new promise that the controversies surrounding the ancestry of the Chinese language can eventually be resolved.

The earliest specimens of Chinese date back some 3,400 years, in the form of inscriptions on animal bones and on bronze artifacts. These inscriptions were made for the purpose of divination or for various historical or ritual purposes. Although both the inscribed characters and the language they represented have changed over the centuries, their ancestral relations with modern characters and contemporary Chinese can be clearly established. Thus of all the modern languages, Chinese can boast of the longest period of continuous use.

Since the characters often contain a phonetic component, they also gave valuable indirect clues for the reconstruction of the spoken language. Efforts to understand the shape, sound, and semantics of the ancient characters have accumulated over the centuries into an impressive body of philological scholarship in China. This scholarship received a major impetus from abroad some 2000 years ago, when Buddhism first came into China. With the religious teachings came the Sanskrit language and its quasialphabetic spelling system, which in turn stimulated the phonetic analysis of the Chinese language.

Under this influence, Chinese scholars divided into the syllable for the time in terms of the fanqie system. A character X can be spelled by two other characters, A and B, where A indicates the initial consonant of X and B indicates the remainder of the syllable of X. The fanqie system fails to help when the reader does not know either A or B. Around 1,700 years ago, Chinese scholars first began writing about the four tones that their language had, which were named *ping, shang, qu,* and *ru*. Rime dictionaries were compiled around 1,500 years ago, phonetic charts were constructed 1,000 years ago, and by 400 years ago, scholars had achieved a reasonable understanding of the history of the language.

Chinese and western traditions began to come together around the beginning of the twentieth century. In 1898, the first western style grammar of Chinese was published by Ma Jianzhong, who studied several European languages while in Shanghai, and who later received training in Paris. Early in the twentieth century, the Swedish scholar Bernhard Karlgren traveled to China to conduct fieldwork on Chinese dialects, which he published as part of his dissertation, 1915–1926. Karlgren was the first to integrate the philological achievements of the Chinese scholars with the real sounds of modern dialects to arrive at systematic reconstructions. Although many aspects of these reconstructions have been superseded, his work has provided an invaluable stimulus for integrating western ideas into Chinese scholarship.

A very important source of data for reconstructions, investigated by Chinese scholars over many centuries, is the Shijing, a body of some 300 poems and songs that date back to around 3,000 years ago. This work is variously known in English as the Book of Poetry, the Book of Songs, or the Book of Odes. By investigating which sets of characters rimed in the Shijing, and by integrating this knowledge with the phonetic information contained in the individual characters, we can go a long way toward inferring how they were pronounced. The language of the Shijing is referred to as Old Chinese. Thanks to the early rime dictionaries, the best known of which is the Qiejun, we can also reconstruct the language of the seventh century, which is referred to as Middle Chinese.

Mandarin can be traced back to another famous rime dictionary, the Zhongyuan Yinyun, compiled in the early fourteenth century. Another important source of information on the sounds of this period is the

'Phags-pa alphabet, invented by a Tibetan lama in 1260, by order of Kubli Khan. The language revealed by these sources is called Old Mandarin.

Phonology

Some of the distinct aspects of Mandarin can be discussed by reference to the structure of the syllable, as shown in the following chart:

Tone				
Initial	Final			
	Medial	Rime		
		Nucleus		Ending

All the dialects of Chinese can be described according to the above chart, since they all share the same syllable structure. The three basic building blocks are the tone, the initial, and the final. In European languages, a syllable is distinguished from other syllables by its vowels, such as 'me' vs. 'moo', or by its consonants, such as 'me' vs. 'bee'. In a tone language, a syllable is distinguished along an additional dimension, as illustrated in the illustration below. A Chinese syllable is also distinguished by its vowels and consonants.

These are illustrated for Mandarin as spoken in Beijing in Figure 1 by 'ma' vs. 'mi' for vowel distinctions, and by 'ma' vs. 'pa' for consonant distinctions. Additionally, 'ma' can carry any of four distinct tones, yielding four completely unrelated words with the meanings of 'to scold', 'hemp', 'mother', and 'horse'. Tone is produced primarily by shaping the pitch contour of the voice by varying the rate of vibration of the vocal folds. The pitch contours of these four words, as pronounced by the present author, are shown in the computer traces in Figure 1. We can see, for example, that the pitch contour for 'to scold' is primarily falling while for 'hemp' it is rising. For other varieties of Mandarin, both the number of tones and the shapes of their pitch contours will be different.

Going back to the chart of the syllable, the initial represents at most one consonant. Thus, there are no consonant clusters in Chinese, such as in English 'spy' or 'pry'. The medial may be any one of the three onglides: 'i', 'u', or 'ü'. The nucleus in Mandarin is a simple vowel, although it may be a diphthong in other dialects. The ending may be any one of the three consonants 'r', 'n', or 'ng'. Thus, a Chinese syllable 'liàng', which means 'light', contains each of the following elements:

Tone: high falling,
Initial: l,
Medial: i,
Nucleus: a,
Ending: ng.

Morphology

In Chinese, by far the great majority of morphemes are represented by single syllables. Since single syllables are written by single characters, there is good correspondence between the minimal units of the sound system, the writing system, and the grammatical system, i.e. syllable = morpheme = character.

In contrast to European languages, word formation in Chinese does not involve any inflectional morphology to speak of, such as one finds in 'books', 'walked', etc. This poses no problem, of course, since information regarding gender, number, and case in nouns and tense in verbs is usually provided by the context. The speaker is not forced by the grammar, for instance, to choose 'he' or 'she' when referring to a friend, since both of these are *ta* in Chinese. On the other hand, Chinese does make use of derivational morphology, such as:

Prefixes— *di wu* 'the fifth', *tou wu* 'the first five',
Suffixes— *zhuantou* 'brick', *zhitou* 'finger', *haizi* 'child', *maozi* 'hat'.

There is an interesting compounding process, whereby opposites are conjoined to form a word. For example, *da* 'big' and *xiao* 'small' are conjoined to mean 'size; *lai* 'come' and *wang* 'go' combine to mean 'social visitation'. A particularly interesting word formed by this process is from *dong* 'east' and *xi* 'west'; the word *dongxi* now has the very general meaning of 'thing' and is used very productively. While English has similar words like 'flip-flop' and 'zig-zag', the two processes are different in that the two members of the Chinese compounding are not phonetically related.

Another interesting process of derivational morphology is reduplication. The process works in several different ways. One area of reduplication has to do with the formation of kinship terms. Whereas most languages have 'papa' and 'mama', such terms are much more abundant in Chinese, e.g. *didi* 'younger brother', *nainai* 'grandmother', *jiujiu* 'maternal uncle', etc. Another area of reduplication has to do with certain nouns to convey the meaning of 'every', e.g. *tiantian* 'every day', *renren* 'every person', *jiajia* 'every family', etc.

Verbs reduplicate, to convey a range of meanings, such as transitory action, e.g. *kankan* 'to take a look', *changchang* 'to have a taste', *tantan* 'to have a chat', etc. When the verb is disyllabic, the reduplication applies to the two syllables together, e.g *shangliang-shangliang* 'to discuss a little', or *yanjiuyanjiu* 'to investigate a little'. Adjectives also reduplicate to form adverbs by taking the particle *de*. Thus, *kuai* means 'quick' and *kuaikuaide pao* means 'run quickly'; *qing*

means 'light' and *qingqingde fangxia* means 'put down lightly'. However, when the adjective is disyllabic, the pattern of reduplication is different from that of the verb. The reduplication here is syllable by syllable, e.g. *an'anjingjingde* from *anjing* means 'quietly' and *gaogaoxingxingde* from *gaoxing* means 'happily'.

In addition to the compounding and reduplication discussed above, another distinctive aspect of Chinese morphology is the use of classifiers. In English, one uses measure words to quantify mass nouns, in expressions such as 'a grain of sand', 'a head of cattle', or 'a piece of cloth'. Many classifiers in Chinese, however, have very little semantic content to speak of, such as in *yige ren* 'one-classifier person', *yipi ma* 'one-classifier horse', *yijian dayi* 'one-classifier coat', *yiben shu* 'one-classifier book', etc. In learning how to use a noun in Chinese, one needs to learn the appropriate classifier for it.

Syntax

Since the language has very little inflectional morphology, grammatical relations among the words are primarily indicated by word order. The basic syntax in Chinese is subject–verb–object, much as in English. A typical declarative sentence with the perfective aspect marker *le* would be:

[1] *Ta mai-le shu.* 'He sell-aspect book.'

A major feature of Chinese syntax is that modificational structures precede the modified. This means that the head is always at the end of phrases. We have seen above that adverbs precede the verb in the verb phrase *kuaikuaide pao* 'run quickly'. In noun phrases, numerals, possessive, adjectives all precede the noun, such as *tade sanben hong shu* 'his three red books'. This applies to relative clauses as well, with the help of the particle *de*, as in *zuotian zai xianggang kande shu* 'yesterday in Hong Kong read-particle book'.

There are two syntactic variants of a declarative sentence: the disposal form with the particle *ba* shown in [2], and the passive form with the particle *bei* shown in [3]:

[2] *Ta ba shu mai-le.*
[3] *Shu bei ta mai-le.*

In these variants, the object is fronted to precede the verb. The meaning of the passive sentence [3] is comparable to that of passive sentences in English. However, the disposal form has no ready counterpart in the syntax of European languages. Although it is a form that is frequently used, its meaning and range of use are topics for which there is no general agreement as yet among Chinese grammarians.

In forming yes–no questions, the typical process for European languages is the inversion of word order, such as 'can he sell books?' formed from 'he can sell books'. English is unusual in the respect that a periphrastic 'do' is used to carry the tense, such as 'did he sell books?' formed from 'he sold books'. Rather than inversion, the typical process for Chinese is to conjoin an affirmative verb phrase with its negative counterpart. The negative counterpart to [1] is [4]:

[4] *Ta mei-you mai shu.* 'He neg-aspect sell book'

Conjoining [1] and [4], we obtain either the question forms [5] or [6] below, depending on whether we delete repeated materials from the affirmative or negative verb phrase:

[5] *Ta mai-le shu mei-you?* 'He sell-aspect book neg-aspect?'
[6] *Ta you mei-you mai shu?* 'He aspect neg-aspect sell book?'

Corresponding to the WH question words in English, i.e. 'who, what, when', Chinese has *shei, shemme,* and *shemme shihou,* respectively. These have been called SH words. In English, WH words need to be moved to the front of the sentence; in Chinese there is no such movement. Going back to [1], we may form the questions [7] and [8] as follows:

[7] *Shei mai-le shu?* 'Who sell-aspect book?'
[8] *Ta mai-li shemme?* 'He sell-aspect what?'

Because of the required movement, the English translation of [8] would be 'What did he sell?'

Every language has a stock of prefabricated phrases that have a richer cultural content, sometimes by means of their historical allusions. Such phrases are learned as complete units, rather than formed spontaneously for the moment. Thus, in English, we might say 'He is the early bird that caught the worm this time'; or 'Aren't you the hostess with the mostess?' Chinese is extremely rich in this regard, particularly in phrases formed with four syllables. These range from colloquial expressions such as *you tou hua nao* 'oily head slippery brain', used to describe someone devious, to forms that are more literary such as *pao zhuan yin yu* 'throw brick induce jade', used as a gesture of humility to say that what you have to offer is not worthy of what you are about to receive.

Because they can have a wide range of cultural associations, prefabricated phrases are sometimes difficult to understand, especially when they are used as part of language games. A striking case of misunderstanding took place when the American writer Edgar Snow interviewed Mao Zedong. Mao described himself with the expression *heshang da shan*, which literally means

'monk with umbrella'. This led Snow to evoke a poignant image: 'He was, he said, only a lone monk walking in the world with a leaky umbrella' [Life Magazine, April 4, 1971]. In actuality, the second part of this four-syllable expression is *wu fa wu tian*, which could mean 'no hair no sky' or 'no law no heaven.' The concept 'heaven' is especially important in this context because in traditional Chinese culture, the ruler governs by *tianming*, i.e. 'mandate of heaven'. The key to this expression is the fact that *fa* is a homonym that means either 'hair', which a monk lacks, or 'law'. The true message Mao was conveying is one of bravado—defying heaven—which Snow completely missed because of the cultural and linguistic gap.

There are many other interesting aspects of the Chinese language: Boltz (1994) gives a good survey of the Chinese writing system, for example, and Huang et al. (1996) discuss the special challenges that Chinese poses to natural language processing—interfacing the language with the computer.

References

Baxter, William H. 1992. A handbook of old Chinese phonology. The Hague: Mouton.

Boltz, William G. 1994. The origin and early development of the Chinese writing system. American Oriental Society.

Chao, Y.R. 1968. A grammar of spoken Chinese. Berkeley, CA: University of California Press.

Huang, Churen, et al. (eds.) 1996. Readings in Chinese natural language processing. Journal of Chinese Linguistics Monograph 9.

Huang, James C.-T., and A. Y.-H. Li (eds.) 1996. New horizons in Chinese linguistics. Dordrecht: Kluwer.

Karlgren, Bernhard. 1954. Compendium of phonetics in ancient and archaic Chinese. Bulletin of the Museum of Far Eastern Antiquities 26. 211–367.

Li, F.K. 1973. Languages and dialects of China. Journal of Chinese Linguistics 1. 1–13.

Masini, Federico. 1993. The formation of modern Chinese lexicon and its evolution toward a national language. Journal of Chinese Linguistics Monograph 6.

Norman, Jerry. 1988. Chinese. Cambridge.

Peyraube, Alain. 1996. Recent issues in Chinese historical syntax, ed. by Huang and Li, 161–214.

Sun, Chaofen (ed.) 1997. Studies in the history of Chinese syntax. Journal of Chinese Linguistics Monograph 10.

Tzeng, O.J.L., and William S.-Y.Wang. 1983. The first two R's. American Scientist 71. 238–43.

Wang, William S.-Y. 1973. The Chinese language. Scientific American, February issue.

Wang, William S.-Y. (ed.) 1991. Languages and dialects of China. Journal of Chinese Linguistics Monograph 3.

Wang, William S.-Y. (ed.) 1995. The ancestry of the Chinese language. Journal of Chinese Linguistics Monograph 8.

WILLIAM S.-Y. WANG

See also **Austronesian; China**

Chinese and Japanese Traditional Grammar

The Chinese linguistic tradition bequeathed an analysis of tone, foreshadowed twentieth-century views of syllable structure, and contributed a moral and political perspective on the question of discrepancy between word meaning and word usage. In addition, explication of the formation of Chinese graphs made an important contribution to the understanding of writing systems. The Japanese tradition, on the other hand, offers unique perspectives on defining parts of speech, and analyzing the ways in which the parts combine to form larger wholes. Both traditions bear a debt to the linguistic tradition of India, some of which was transmitted through the powerful vehicle of Buddhism, in the intersection of phonetic analysis and sacred sound. Also common to both was the impetus of early poetry, in the desire to know the precise sounds and meanings of the past.

The Chinese debate on 'rectification of names' originates in several passages from the *Analects* of Confucius, wherein the belief that use of words should reflect reality is asserted. When it does not, the people cannot be properly governed. Xun Zi (c. 313–238 BCE), regarded as China's first major philosopher of language, contributed to this debate in a work titled *Zheng Ming* (*Rectification of names*), which addresses the origin of names, in terms similar to those found in Plato's *Cratylus*. Unlike *Cratylus*, however, the main focus does not concern whether language is to be trusted as epistemology. In Xun Zi's view, names are established by convention, and have no inherent correctness. They thus cannot represent absolute standards for human behavior.

Hsü Shen (58–147 CE) compiled the *Shuo wen jie zi* (first century CE, *Explanations of simple graphs and*

analyses of composite graphs), the first major lexicon, and one of the most significant works of the Chinese tradition. The 9,353 graphs contained in the work are organized under 540 radicals. One reason for its landmark status derives from the fact that this is the first surviving account of the six principles of formation of the Chinese graph, including that of semantic radical plus phonetic element, which accounts for more than 90% of modern graphs.

A poet, Shen Yue (441–513), is traditionally credited with the first analysis of Chinese tones, recognizing and naming four types. The tradition of rhyme books is thought to have originated as an aid to composition of poetry, and the technique of *fan qie* (reverse cutting) used therein prefigures twentieth-century notions of the syllable as composed of an internal structure, with onset and rhyme. The purpose of the rhyme books was to indicate the pronunciation of the monosyllabic graphs of Chinese, and did so by first making four major tonal divisions, and then within each tonal section, listing graphs in charts arranged by initial sound, and by the remainder, or rhyme. The so-called 'cutting' involved indicating pronunciation of a given graph by providing two other graphs having the same initial and rhyme portions, respectively. The *Qie yun* (601 CE), compiled by Lu Fayan, is the earliest major example of the genre.

Rhyme tables, the earliest known example of which is the *Yunjing* (twelfth century, *Mirror of rhymes*), provided a more detailed classification of all of the syllables of Chinese. In these works, initials are arranged according to articulatory categories such as 'tongue sounds', or dentals, and characterized in terms of phonation by such terms as 'clear' (voiceless) or 'muddy' (voiced).

In Japan, early linguistic studies were concerned with determining the sound correlates to different sets of phonograms in Old Japanese, which had merged into single sounds in later stages of Japanese. Notable in their efforts in this regard were the poet Fujiwara Teika (1162–1241) and Buddhist monk Keichû (1640–1701).

Analysis of morphology and syntax in traditional Japanese grammar is largely a concern with tactics: the combinatorial properties and functions of various verbal and adjectival suffixes, on the one hand, and postpositions or particles, on the other. Fujitani Nariakira (1738–1779) created the first comprehensive grammatical framework, using the metaphor of clothing: he categorized nouns as 'names', verbs and adjectives as 'clothes', adverbs as 'hairpins', and postpositions and suffixes as 'binding cords'. His best-known works are the *Kazashishô* (1767, *On hairpins*) and the *Ayuishô* (1773, *On binding cords*).

In his *Teniwoha himokagami* (1771, *Survey of particles*), Motoori Norinaga (1730–1801) explicated the ways in which certain emotive particles trigger concord with inflectional endings in the verb, and coined the term *kakari-musubi*, still in use today, to refer to it. In *Mikuni kotoba katsuyôshô* (1782, *On inflections in our language*), Motoori laid the groundwork for understanding the patterns of verbal conjugation.

Suzuki Akira (1764–1837), student of Motoori Norinaga, in his *Gengo shishûron* (1824, *On the four categories in language*), breaks down parts of speech into nouns, verbs (activities), and adjectives (states), on the one hand, and particles and bound suffixes, on the other. Underlying this categorization is a conceptualization of content vs. function words, said to have been drawn under the influence of such a classification for Chinese. These grammatical studies from the early modern period fall under the scholarly domain of *kokugaku* (nativism), while the work of those continuing in the same vein in the twentieth century is termed *kokugogaku* (national language studies), as separate from the western-influenced discipline of linguistics (*gengogaku*).

The most influential twentieth-century figure to continue along nativist lines was Tokieda Motoki (1900–1967), who expanded the framework to encompass all languages, without restriction to Japanese. He added a subjective, performative dimension to the bifurcation of content words, or *shi* (in his view, objectively conceptualized) vs. function words, or *ji*, which express the emotions, attitude, or judgments of the speaker. Under this view, content words are nested within function words in an iconic display, which leaves the subjective elements rightmost at each level of grammatical construction.

Major work on Japanese dialects was carried out by the folklorist Yanagita Kunio (1875–1962) in his *Kagyû-kô* (1927, *Thoughts on 'Snail'*), where he posited the 'circle theory' of linguistic diffusion, and linguist Hattori Shirô (1908–), in his investigations of the geographical distribution of accent patterns. The National Language Research Institute has published two major dialect atlases: the *Linguistic atlas of Japan* (1981–1985) and the *Grammar atlas of Japanese dialects* (1989–).

References

Baxter, William H. 1992. A handbook of old Chinese phonology. Berlin and New York: Mouton de Gruyter.

Chaudhuri, Saroj Kumar. 1998. Siddham in China and Japan. Sino-Platonic Papers, No. 88.

Furuta, Tôsaku, and Hiroshi Tsukishima. 1972. Kokugogaku shi (History of Japanese language studies). Tokyo: Tokyo Daigaku Shuppan Kai.

Hansen, Chad. 1983. Language and logic in ancient China. Ann Arbor: The University of Michigan Press.

Malmqvist, Göran. 1994. Chinese linguistics. History of linguistics, Vol.1: The Eastern traditions of linguistics, ed. by Giulio Lepschy. London and New York: Longman.

Miller, Roy Andrew. 1975. The Far East. Current trends in linguistics, Vol.13: Historiography of linguistics, Part 2, ed. by Thomas A. Sebeok. The Hague and Paris: Mouton.

Motoori, Norinaga. 1997. Kojiki-den, Book 1, translated by Ann Wehmeyer. Ithaca, New York: Cornell East Asia Series.

Ramsey, S. Robert. 1982. Language change in Japan and the odyssey of a Teisetsu. Journal of Japanese Studies 8.

Ramsey, S. Robert. 1987. The languages of China. Princeton, NJ: Princeton University Press.

Sakai, Naoki. 1992. Voices of the past: the status of language in eighteenth-century Japanese discourse. Ithaca, New York: Cornell University Press.

ANN WEHMEYER

See also **China; Chinese (Mandarin); Japanese**

Chinese Pidgin English

Were Chinese Pidgin English not widely known by that name, it would be preferable to use the alternative, China Coast Pidgin. While having significant influence, English alone did not provide the core of conventions that came to constitute the language. Instead, there was convergence of English and Cantonese features in a process of mutual accommodation. Indeed, the hybrid nature of the structures of Chinese Pidgin English makes it a crucial piece of evidence in debate within creole linguistics, reflecting a shift away from particularist notions of pidgins and creoles to one in which they are considered merely relative, although related, forms of contact languages.

Chinese Pidgin English originated during the first locally regulated phase of trade between Chinese and English in the late seventeenth century (c. 1689–1748). It found a stable form in and around Canton for about another century (1748–1842), and was thereafter disseminated further up the Chinese coast during the remainder of the nineteenth century, and beyond by the Chinese diaspora to the Pacific and even the United States (1842–1890). Finally, Chinese Pidgin English fell into obsolescence during the twentieth century when it was heavily anglicized and lost many of its original structures as numerous speakers came to target English as a lingua franca.

Unfortunately, with the exception of slender second-hand evidence from the first period of genesis, mostly in European travel accounts, almost all the data we have fall between 1836 and 1901, and, heavily anglicized, are not reliable. The early data, however tenuous, nonetheless reflect a process of mutual accommodation in which speakers had to make guesses about what their interlocutors would understand and 'right' guesses would be incorporated into the grammar of the developing contact language. Balanced against the predominance of English etymons in Chinese Pidgin English is the relative importance of Cantonese syntax and phonology, such that to most English speakers, Chinese Pidgin English was hardly more intelligible than Chinese itself. There is convincing documentary evidence that the right guesses that set the parameters of Chinese Pidgin English took place within the trade factories at Canton between 1699 and 1748, where its inchoate forms were the instrument of limited exchange between the 'supercargoes' or British East India Company property agents, and the Cantonese-speaking domestics assigned to them by their authorized 'Hong merchant' interlocutors.

One objection to such a scenario is the undeniable impact of Portuguese on Chinese Pidgin English, one that might lead us to suspect that its origins go back to the arrival of the Portuguese along the south China coast in 1557—English 'Canton' is derived from the Portuguese pronunciation of *Guangzhou*. Some of the earliest fragments of putative Chinese Pidgin English appear to contain more Portuguese than English, which would tend to point toward a Portuguese base, e.g. this phrase from 1748:

(1) Carei grandi hola,   pickenini hola?
    Want   large   whore, small   whore
    Quer   grande [puta], pequena   [puta] (Ptg)
    'Do you prefer a big or a small prostitute?'

There is no doubt that Portuguese or some pidginized form of it was spoken along the China coast up to and throughout the time Chinese Pidgin English was formed. In Macao, there eventually formed *Macaista*, a variety of creolized Portuguese. Maritime and pidginized Portuguese were used from the Indian subcontinent across through Malaysia and Indonesia. As for Canton itself, we know that the first requirement of a supercargo on English ships trading to China was a knowledge of Portuguese. Yet, in the final analysis it is unlikely that a prior Portuguese pidgin served as the base of incipient

Chinese Pidgin English, not so much because the Portuguese contribution was limited to an ever-decreasing percentage of its lexicon but rather because the role of Cantonese is predominant both in pronunciation and word structure, from which fact we can suppose that the majority of early pidgin learners were native Cantonese speakers who adopted the vocabulary of the supercargo masters to whom they were assigned.

The Canton-based pattern of trade that formed Chinese Pidgin English continued until the end of the eighteenth century. In 1796, the Emperor banned the opium trade, at which point the English turned to smuggling. In 1840, a second attempt was made to ban opium, which led to the Opium War and the Nanking Treaty of 1842, opening the doors to China. The focus of contact shifted from Canton to the Treaty cities and to Hong Kong. As the monopoly of the Hong Kong merchants broke down, contacts began to proliferate between aspiring Chinese entrepreneurs and the quickly increasing number of foreigners in south China. The habit of resorting to Chinese Pidgin English was carried into new sites of trade where the status and practice of the language changed. As speakers of standard English began entering the Chinese community from the late nineteenth century onward, Chinese Pidgin English became increasingly associated with the servants of foreigners, and was deprecated. Eventually, Chinese Pidgin English was conveyed first as far north along the coast as Shanghai and Tientsin, British

traders taking their Cantonese servants with them as China was forced open to trade. After economic conditions worsened throughout the nineteenth century, there was a large Cantonese diaspora to California, Hawaii, Australia, and Singapore. At the same time, the core of Chinese Pidgin English, learned and transmitted as an auxiliary second language for several generations, began to dissolve. It was this late period of its history, roughly from 1842 onward, that gave rise to stereotypes and misunderstandings, which have tended not only to mask the originality of this language but also to confuse it with all forms of imperfectly learned English, which it was not.

References

Bauer, A. 1975. Das Kanton-Englisch: ein Pidginidiom als Beispiel für ein Kulturkontaktphänomen. Bern: Peter Lang.

Holm, John. 1988–1989. Pidgins and Creoles, 2 vols. Cambridge: Cambridge University Press.

Lang, George. 2000. Hardly more intelligible than Chinese itself: a brief account of Chinese Pidgin english. Asian Englishes 3(1). 21–38.

Morse, Hosea Ballou. 1966. The chronicles of the East India Company trading to China, 1635–1834. Taipei: Ch'eng-Wen Publishing Company.

Shi, Deng-Xi. 1991. Chinese Pidgin English: its origins and linguistic features. Journal of Chinese Linguistics 19. 1–40.

Thomason, Sarah Grey, and Terrence Kaufman. 1988. Language contact, creolization, and genetic linguistics. Berkeley: University of California Press.

<div align="right">GEORGE LANG</div>

Chinese Pidgin Russian

Not surprisingly, the Russian expansion across the Ural mountains into Siberia and America led to a number of contact languages to a greater or lesser extent lexically based on Russian. We have documentation of such varieties from four locations—Kyakhta (on the Mongolian border), Manchuria, the Taymyr Peninsula, and in the Ussuri area just north of Machuria. The respective pidgins are hereafter abbreviated as KPR, MPR, TPR, and UPR.

It is also possible that further varieties exist or have existed in Siberia, as well as in areas of (attempted) Russian colonization in Alaska, California, and Hawaii. In addition, mention will be made of Russenorsk (RN), a trade pidgin of the Arctic Ocean based partly on Russian and partly on Norwegian. I will not, however, discuss Mednyj Aleut, spoken in

the Strait of Bering, since, although a contact language, it is not a pidgin (but rather a variety of Aleut with Russian verbal morphology).

MPR emerged in the very beginning of the twentieth century, when Russians managed the Manchurian railway. It began to decline after World War II, when both parties took an increasing interest in learning the other's language, and definitively dropped out of use when the new political climate put an end to Sino-Russian friendship. People who remember the pidgin are still alive, however; thus, research is still possible, but urgent.

KPR emerged as a result of trade between Russians and Chinese in the eighteenth century. The pidgin has been considered extinct, but in 1990 some speakers were found among Chinese merchants in the Mongolian capital of Ulan Bator.

TPR, also called *Govorka*, is used between speakers of Samoyedic, Turkic, and Tungusic languages on the Taymyr Peninsula in northernmost Siberia. It has been declining in favor of Russian since World War II, and is now spoken only by elderly people. Fortunately, research on TPR is currently carried out by German linguist Dieter Stern.

UPR is the least well documented of the Russian pidgins, at it is known virtually exclusively through accounts of Russian expeditions to the area in the late nineteenth and early twentieth centuries.

Russenorsk, finally, was used in the trade between Russians and Norwegians in the border area until the beginning of the twentieth century. There is a relative wealth of material on this pidgin, which has been analyzed primarily by Norwegian linguists.

Although Russenorsk and TPR are no doubt local developments, it should be emphasized that the relationship between the Far Eastern varieties is unclear, and, indeed, many authors treat them all as 'Chinese Pidgin Russian'.

In most of the pidgins, there is a tendency toward syllabic simplification, favoring CV syllables. Other phonological aspects are highly variable, as they display influences from surrounding languages, such as the devoicing of Russian plosives in MPR, the replacement of /f/ with /p/ in KPR, and the merger of /s, z, ʃ, t͡ʃ/ in TPR.

Lexically, some of the varieties are remarkably mixed. This is particularly true for RN, in which only about half of the vocabulary was of Russian origin, and MPR, where approximately one third of the lexicon was non-Russian (mainly from Mandarin). Some lexical influence from local languages is reported for the other varieties as well. In addition to the lexical items themselves, some calquing from substrate languages has also been observed.

As in all pidgins, the small lexicon entails an exceptional degree of polysemy and multifunctionality, the serious semantic underspecification leading to a strong reliance on context for disambiguation.

As for the derivation of the Russian forms, it is noteworthy that the verbs are often developed from imperatives rather than from infinitives, as in pidgins lexically based on Western European languages.

The basic word order is of particular interest, for only MPR was consistently SVO, like both Russian and Mandarin. RN was basically SVO, but used SOV with adverbials. The other varieties are predominantly SOV, presumably due to substrate influences. TPR is also spectacular in having the order REL N. Another un-Indo-European feature is the optional lack of fronting of interrogative pronouns. As would be expected from SOV languages, postverbal auxiliaries are attested.

Tense/Mode/Aspect (TMA) marking is usually rather restricted. TPR encodes tense, but not aspect. If the little data available there are indeed reliable, UPR is remarkable, and possibly unique among pidgin langauges, in having a grammaticalized evidentiality marker. This, however, seems to have been the only TMA category in the language. Russenorsk had no grammaticalized TMA marking at all, whereas the other Far Eastern varieties seem to have made optional use of nonbound morphemes for this purpose.

Negation is Russian style in MPR, but postverbal in the other Asian varieties. Russenorsk had a preverbal negation, or rather two of them, which usually occupied the second position in the sentence. Nominal number is usually left unmarked, although TPR has optional morphemes derived from Russian words meaning 'much' and 'they all'. Grammatical gender has not survived in any of the varieties.

One of the shibboleths of pidginhood is a limited adpositional inventory. Russenorsk had *po* as its only preposition, corresponding for the most part to *za* in the Far East. These items expressed more or less any spatial relationship imaginable. The same job is carried out by *mesto* in TPR, which, however, also has a sociative adposition *meste*. It is noteworthy that adpositions in TPR are post- rather than prenominal.

Reduplication is absent from Russenorsk and MPR, but has been reported for KPR.

As in many other pidgins, juxtaposition is frequently used to indicate possessive relationship, and in a similar manner, clauses are to a great extent joined paratactically without any overt conjunction or subjunction (although such clauses do exist).

Again, as one would expect from pidgins, bound morphology is scarce, although there are a couple of potential candidates for the status of bound morphemes.

The pronominal systems also merit mention. The forms themselves, invariable with regard to syntactic function, are mostly derived from Russian genitives, possibly a carryover from Russian foreigner talk. Exceptions to this are the TPR forms, and the 2sg form elsewhere, which derive from Russian accusatives. They have been subject to some rather far-reaching restructuring, however, in particular in TPR. There, only the singular forms have been taken over, with additional morphemes marking number. Even more conspicuous is the presence of an inclusive/exclusive distinction for nonsingular pronouns.

References

Jabłonska, Alina, and Anatole Lyovin. 1969. The Sino-Russian mixed language in Manchuria. University of Hawaii Working Papers in Linguistics 1. 135–64.

Neumann, Günther. 1966. Zur chinesisch-russischen Behelfssprache von Kjachta. Die Sprache 12. 237–51.

Nichols, Johanna. 1980. Pidginization and foreigner talk: Chinese Pidgin Russian. Papers from the 4th International Conference on Historical Linguistics, ed. by Elizabeth Closs Traugott, Rebecca Labrum, Susan Shepherd, and Paul Kiparsky, 397–407. Amsterdam: Benjamins.

Nichols, Johanna. 1986. The bottom line: Chinese Pidgin Russian. Evidentiality: The linguistic coding of epistemology, ed. by Wallace Chafe and Johanna Nichols, 239–57. Norwood: Ablex.

Peter Mühlhäusler, Darrell T. Tryon, and Stephen A. Wurm (eds.) 1997. Atlas of languages of intercultural communication in the Pacific, Asia, and the Americas. Berlin and New York: Mouton de Gruyter.

Wurm, Stephen. 1992. Some contact languages and Pidgin and Creole languages in the Siberian region. Language Sciences 14(3). 249–85.

Wurm, Stephen. 1996. The Taimyr Peninsula Russian-based pidgin. Language contact in the Arctic. Northern Pidgins and contact languages, ed. by Ernst Håkon Jahr and Ingvild Broch, 79–90. Berlin and New York: Mouton de Gruyter.

Xelimskij, Evgenij. 1996. 'Govorka' — the pidgin Russian of the Taymyr Peninsular Area, In Mühlhäusler Tryon and Wurm, pp. 1033–4.

MIKAEL PARKVALL

Chinook Jargon

Chinook Jargon is a pidgin language of the Pacific Northwest that is first attested reliably from the first decade of the nineteenth century, in the journals of Lewis and Clark; the earliest extensive documentation is by Horatio Hale (1846). Its lexifier language—the language from which most of its vocabulary is drawn—was Lower Chinook (Shoalwater), the language of a once-powerful tribe at the mouth of the Columbia River. At its peak, Chinook Jargon was spoken as far north as southern Alaska, as far south as the northern border of California, and as far east as the Idaho panhandle in the United States and interior British Columbia in Canada. It flourished especially between c. 1850 and 1950, when it was a primary medium of communication between Whites, on the one hand, and Native Americans and Native Canadians on the other. Its use declined sharply in the late twentieth century as English replaced it everywhere in the Northwest. As late as 1980, however, monthly sermons were delivered in Chinook Jargon in at least one British Columbian church, and a few elderly fluent speakers on the Grand Ronde Reservation in Oregon contributed to its survival by teaching it to younger tribal members. Efforts are currently under way to revitalize Chinook Jargon, especially at Grand Ronde.

By the late twentieth century, Chinook Jargon was the main or only Native language spoken on the reservation, where members of six Native tribes descended from speakers of three completely different indigenous language families. As Henry Zenk has noted (1984), Chinook Jargon became 'an important factor in the sense of identity and solidarity that many Natives of the reservation period came to feel as "Grand Ronde Indians".' Chinook Jargon thus performed its traditional function as a lingua franca even toward the end of its continuous existence.

Like other classic trade pidgins, Chinook Jargon has a limited vocabulary. Over 600 Chinook Jargon words are reliably attested—that is, they occur in at least two independent sources, and usually in more than two. Of these, perhaps a third come from French or English, reflecting the widespread use of Chinook Jargon as a lingua franca between Whites and Natives in the late nineteenth and early twentieth century. The earliest Chinook Jargon sources, from the first half of the nineteenth century, also list English and French words, but not nearly as many as in the later documentation. A rough count of c. 650 Chinook Jargon words must certainly omit many that were in common use: even the most ambitious dictionaries inevitably omit many of a language's words, and the published Chinook Jargon dictionaries (e.g. Shaw 1909) are not especially ambitious.

Chinook Jargon phonemes are typical for an indigenous language of the Pacific Northwest. More than 30 consonant phonemes are attested, in a few to numerous words, in at least two independent sources each. Many of the consonants are unfamiliar to speakers of English and other European languages: glottalized (specifically, ejective) stops and affricates, uvular as well as velar obstruents, a glottalized lateral affricate, and a lateral fricative. Moreover, Chinook Jargon words often contain consonant clusters that are completely foreign to European languages, e.g. in /tk'up/ 'white, light in color', /ptSix/ 'thin', and /tL'm@n/ 'soft, ground up'. Chinook Jargon thus presents a striking counterexample to the often-repeated claim that pidgin structures are maximally simple, and the consistency of many of its phonological features

across Native speakers from numerous tribes offers a counterexample to the common claim that pidgin structures show more internal variation than nonpidgin languages. To judge by the surviving documentation and the few existing tape recordings from the mid-twentieth century, most White speakers of Chinook Jargon did not learn either the non-European sounds or the non-European consonant clusters of the pidgin (but see Demers et al. (1871) for a notable exception).

Also, like other classic trade pidgins, Chinook Jargon has limited morphosyntactic resources. It entirely lacks the complex morphological (word) structures that characterize Native languages in the region, and its range of syntactic constructions is not large. As with the sound system, its syntactic structures closely resemble those of Native Northwest languages—with one possible major exception, the dominant S(ubject)–V(erb)–O(bject) word order. Most Northwest languages are verb-initial; Chinook Jargon syntax is verb-initial only with adjectival predicates, e.g. _Hayas ulu tsuq nayka_ ‘I am very thirsty’ (lit. ‘much hungry water I’). Besides the SVO word order, the constructions are clearly Native, not based on either English or French. They include, among others, sentence-initial negation, yes/no questions formed with an optional question particle, and an imperative construction (‘it would be good if you would do X’).

The question of the pidgin’s origin remains highly controversial. One hypothesis is that Chinook Jargon predates extensive contact with Whites in the Northwest—that it was used as a means of intertribal communication, perhaps at first between speakers of Lower Chinook and their Native slaves. (‘Slave jargons’ are reported elsewhere in the Northwest, for instance among the Nez Perce.) Based on this theory, Chinook Jargon achieved its later spread when Whites adopted it for use as a lingua franca, shortly after 1806. The second origin hypothesis is that Chinook Jargon arose as a lingua franca only after Whites arrived in large numbers in the Northwest. A common feature of this theory is the proposal that a Nootka trade jargon (or pidgin) arose first on Vancouver Island, at the end of the eighteenth century or early in the nineteenth century, and then spread to the mouth of the Columbia River when Whites ventured there for trade; the Nootka trade jargon/pidgin was then relexified through the replacement of most Nootka words by Chinook words.

Evidence adduced in support of the second theory is primarily lexical. Most of the lexicon of early Chinook Jargon, including the bulk of the basic vocabulary, comes from Lower Chinook, but two or three dozen words (some of them quite basic) are from Nootka, and a smaller number of words come from Salishan and other languages of the region. In addi-tion, quite a few words entered the pidgin from French and then English, especially after about 1850.

The hypothesized contribution of the Nootka Jargon to the initial formation of Chinook Jargon is predicated on the assumption that such basic words would necessarily have been in Chinook Jargon from the beginning, not added later after Chinook Jargon was fully formed. Against this assumption, however, is the undoubted fact that, while the Nootka words in Chinook Jargon show clear signs of transmission from Whites, the Chinookan and Salishan words in Chinook Jargon were clearly transmitted from Natives to other Natives. Words of Chinookan and Salishan origin show all the elaborate features of typical Northwest phonological systems, including glottalized stops, velar vs. uvular dorsal obstruents, and lateral fricatives and affricates. This is not the case with Nootka-origin words: they are significantly distorted in comparison to their Nootka source words, with virtually no sounds that would be foreign to English and French speakers. The transmission of some of the Nootka jargon/pidgin words to Chinook Jargon is therefore much more likely to have occurred after, not before, the crystallization of Chinook Jargon as a pidgin language.

Both origin theories are plausible. The postcontact theory has been popular in part because Chinook Jargon, from the time it was first documented, already had the Nootka words and a fair number of French and English words. Also, of course, all Chinook Jargon documentation is necessarily postcontact, since Natives in the Northwest had no writing before Whites arrived. The precontact origin theory is preferable if one adopts the standard simplicity criterion of historical linguistics: the Native phonology and syntax are easily accounted for if Natives created Chinook Jargon without significant participation by Whites, but if the pidgin arose postcontact, with some French- and/or English-influenced structural features, those features must have been lost before the pidgin was documented. Based on this criterion, the precontact origin is the simpler hypothesis. But not all pidgin/creole specialists accept this criterion; hence, the controversy continues.

The future of Chinook Jargon is in some doubt, because the Grand Ronde elders and others who spoke the language as part of their ordinary daily lives are now gone. Its fate rests with the younger enthusiasts, especially younger tribal members, who are now working to prevent the pidgin from disappearing.

References

Demers, Modeste, F. N. Blanchet, and L. N. St. Onge. 1871. Chinook dictionary, catechism, prayers, and hymns. Montreal: Quebec Mission.

Hale, Horatio. 1846. United States exploring expedition during the years 1838–1842; ethnography and philology. Philadelphia: Sherman.

Samarin, William J. 1986. Chinook Jargon and pidgin historiography. Canadian Journal of Anthropology 5. 23–34.

Shaw, George C. 1909. The Chinook Jargon and how to use it. Seattle: Rainier Printing Co., Inc.

Splawn, A.J. 1944. Ka-Mi-Akin: last hero of the Yakimas, 2nd edition. Yakima, WA, Caldwell, Idaho: The Caxton Printers; 1st edition, 1917.

Thomason, Sarah G. 1983. Chinook Jargon in areal and historical context. Language 59. 820–70.

Zenk, Henry Benjamin. 1984. Chinook Jargon and Native cultural persistence in the Grand Ronde Indian community, 1856–1907: a special case of creolization. Eugene, OR: University of Oregon dissertation.

SARAH G. THOMASON

See also **Language: Contact—Overview; Pidgins and Creoles**

Chomsky, Noam

Noam Chomsky is one of the most profound and influential thinkers of our time, 'arguably the most important intellectual alive' (*New York Times Book Review*, February 25, 1979), an evaluation in accord with the fact that he is the world's most cited living author in several citation indexes.

In his early years, Chomsky read Hebrew literature with his father, a professor of Hebrew and Jewish education at Gratz College in Philadelphia (and president of its faculty for 45 years), considered 'one of the world's foremost Hebrew grammarians' (*New York Times* obituary, July 22, 1977). His immersion in the Jewish cultural tradition was deep. Out of this experience, his budding political interests converged toward what was then mainstream Zionism (now widely considered 'anti-Zionism').

Politics brought Chomsky into linguistics. As a teenager, he was deeply interested in radical politics with an anarchist or left-wing Marxist flavor (strongly anti-Leninist, and more generally anti-Bolshevik). Through these political interests, he met Zellig Harris, whom Chomsky has described as 'a person of unusual brilliance and originality' and very broad interests. Harris was not only an acute left-libertarian thinker and analyst but also a leading figure in linguistics, teaching at the University of Pennsylvania, and the most rigorous practitioner of the reigning structuralist methodology. Harris was surely the linguist best prepared to initiate someone with Chomsky's mind and inclinations into the field of linguistics (as then understood). He also interested him in the study of philosophy, logic, and mathematics, fields that were to open for him the avenue to major discoveries. The first reading Chomsky did in linguistics, before he had taken any courses, was the proofreading of Harris's important *Methods of structural linguistics* (1951), the most exacting exposition of immediately pre-Chomskyan linguistic theory. The Preface (dated January 1947) records that 'N. Chomsky has given much-needed assistance with the manuscript.'

Since the appearance of his first published book, *Syntactic Structures*, in 1957, Chomsky has been recognized as 'an eminent and revolutionary scholar in the field of linguistics', to quote from a representative blurb. Chomsky almost single-handedly assimilated psychology to the natural sciences by formulating 'transformational generative grammar' (a model for the 'cognitive sciences', of which it was the first). This theory takes for granted the now widely (although not unanimously) held view that all human languages operate under the same general principles in spite of their superficial diversity. Particularly important in this context is that the universal foundation for language is taken to be inherent in (and an essential part of) 'human nature', this being the central notion in both of Chomsky's main endeavors. Thus, Chomsky rejected behaviorism, unequivocally so in his renowned 1959 review of B.F. Skinner's 1957 *Verbal behavior*, a review that demonstrated the inability of the behaviorist approach to account for any but the most superficial aspects of human language.

Chomsky redefined the nature and scope of linguistics. As the *Encyclopedia Britannica* (1992) says, '... there is no major theoretical issue in linguistics today that is debated in terms other than those in which he has chosen to define it'. A recent book on the future of science describes him, accurately, as 'the most important linguist who has ever lived'. As John Lyons says in *Chomsky* (1970), he has spoken 'with unrivaled authority in all aspects of grammatical theory' since *Syntactic structures* 'revolutionized the scientific study of language', fundamentally changing the then current understanding of language and the mind/brain.

Chomsky's most fundamental contribution has been to open the way for the cognitive natural sciences. His

specific model for human language (transformational generative grammar) can be regarded as a confluence of traditional (and long-forgotten) concerns of the study of language and the mind (as in the work of Wilhelm von Humboldt or Otto Jespersen) and new understanding provided by the formal sciences in the late 1930s, particularly recursive function theory. The basic idea is that a sentence is the result of a computation producing a 'derivation', beginning with an abstract structural representation sequentially altered by structure-dependent operations ('transformations'). Chomsky's unusual knowledge of philosophy, logic, and mathematics allowed him to make this model precise, and to go on to develop, in the late 1950s, algebraic linguistics, a branch of abstract algebra, which is now part of computer science. (For Israeli logician and mathematician Yehoshua Bar-Hillel, one of its most eminent practitioners, Chomsky was not just the founder of algebraic linguistics but also 'by far the best man in this exciting new field'.)

Chomsky's written output is truly prodigious, encompassing, in linguistics alone, dozens of books and nearly 200 articles (and several times that in political analysis). His first substantial contribution to the development of generative grammar was his B.A. thesis (1949), an examination of certain morphophonemic alternations in Modern Hebrew, revised and expanded in his M.A. thesis (1951), and revised once more that year (that version was eventually published in 1979). Although this work had no impact on the field of linguistics at the time, it did introduce numerous technical devices that would eventually be of major significance, including abstract underlying syntactic and phonological forms and crucially ordered rules deriving surface forms from these underlying forms. Significantly, already in this early work, Chomsky was emphasizing that a grammar is a finite characterization of an infinite set of sentences. This theme continues to be of profound importance in linguistics.

Over the next few years, while a Junior Fellow of the Society of Fellows at Harvard University, Chomsky wrote *The logical structure of linguistic theory* (LSLT), a monumental work (completed in 1955, but published, and only in part, 20 years later) that laid out the formal basis for a complete theory of linguistic structure. The concepts and technical notions that were to become central to theoretical linguistics for the next several decades were developed in this paper, including the crucial idea of abstract underlying structure. Made precise are such concepts as level of representation, phrase structure rule, phrase marker, grammatical transformation, and derived constituent structure. LSLT is also rich in conceptual and empirical arguments, and it provides an extraordinarily detailed account of the syntax of English. Many of the patterns Chomsky explained had never even been noted before. Further, many of the analyses he presented have yet to be surpassed. However, like the M.A. thesis, LSLT had no direct impact on the field at the time. Within two years, though, LSLT had a revolutionary indirect influence, since the material contained in it formed the basis for the lecture notes that were published in 1957 as *Syntactic structures*, a book whose unprecedented influence was noted above.

The year 1965 witnessed the publication of Chomsky's next major book, *Aspects of the theory of syntax*, which made fully explicit the role of linguistics in the investigation of the human mind. Its first chapter, originally written during the year Chomsky spent at the Institute for Advanced Studies at Princeton (1958–1959), remains one of the clearest and most forceful expositions of the major goal of generative theorizing: providing an account of how the child presented with limited evidence develops a computational system making possible the production and comprehension of an unlimited number of brand new sentences ('explanatory adequacy'). The book laid out what came to be called the 'standard theory' of syntax. In this theory, the syntactic derivation of a sentence begins with deep structure (DS), a representation constructed via phrase structure rules and lexical insertion rules. Grammatical relations are fully determined at this level. Transformations then successively, and cyclically ('bottom up'), modify this representation, eventually producing surface structure (SS), the input to the phonological component. This contrasts with the LSLT model, wherein the recursive component is the transformational one: ever larger structures are created not by phrase structure rules but by generalized transformations, which embed one structure inside another. (Interestingly, in Chomsky's 'minimalist program' of the 1990s, there is a return to generalized transformations.) Versions of the standard theory (including the 'extended standard theory' and the 'revised extended standard theory') dominated syntactic theorizing for more than two decades.

While Chomsky is most closely associated with syntactic theory, much of his early work was concerned with phonology, culminating in the groundbreaking book in 1968, *The sound pattern of English* (co-authored by Morris Halle), which was a detailed study of the phonology of English, but, just as importantly, a full-blown theory of the phonological component of linguistic theory and its interaction with the syntactic component. Also noteworthy is the elaborated outline of a theory of markedness, which sought to explain why certain phonological processes are much more common than others.

Chomsky's concern with explanatory adequacy led in due course to the 'Government-Binding' model,

articulated in *Lectures on government and binding* (1981). In this model, the language (and dialect)-specific and construction-specific rules of earlier frameworks are replaced by operations of great generality whose functioning is constrained by universal conditions. The rules and conditions are 'principles', and the limited range of variation available for them constitutes the 'parameters'. (Chomsky came to prefer the name 'Principles and Parameters' for the model, reasoning that government and binding are just two among many technical devices in the theory, and not necessarily the most important ones.) The P&P model—the first one ever to suggest a substantive solution to the fundamental problem of language acquisition—represents a radical break from the rich tradition of thousands of years of linguistic inquiry. It is true, though, that the entire (modern) generative grammar period is in many ways a new era.

The P&P model was in turn further refined in Chomsky's 'minimalist program' of the 1990s and early 2000s (see Chomsky (1995), the second chapter of which was written and circulated in 1988, based on lectures in 1986 and 1987). The P&P model had DS, SS, Logical Form (LF), and Phonetic Form (PF) as significant levels of representation. Given that a human language is a way of relating sound (or, more generally, gesture, as in sign languages) and meaning, minimalism seeks to establish that there are no levels except the 'interface' levels PF and LF. In the most recent developments of the theory, Chomsky suggests, in another partial return to an earlier formulation, that even PF and LF, as specific levels of representation in the technical sense, do not exist. Rather, throughout the derivation, the syntactic structure thus far created is encapsulated and sent off to the interface components for phonetic and semantic interpretation. The minimalist program further maintains that derivations and representations conform to an 'economy' criterion demanding that they be minimal in a sense determined by the language faculty: no extra steps in derivations and no extra symbols in representations. A major technical goal is to reduce all constraints on representation to 'bare output conditions', determined by the properties of the mental systems that the linguistic computational system must interface with that are external to the computational system itself.

A direct consequence of Chomsky's scientific work is that it provides strong evidence in support of epistemological rationalism, as he was quick to point out. Early on, he turned to the serious study of the Cartesian tradition, which he was to revive and update, shortly after he made his initial discoveries. These discoveries made it possible for him to go well beyond the programmatic insights of the Cartesians, and give substance to their central claims, in the process reconstructing the enduring ideas of the first phase of the age of modern philosophy (a not always recognized antecedent of the cognitive revolution of the 1950s), on which Chomsky has shed much light as an intellectual historian. It seems fair to describe him as the scholar who is to the period initiated by the cognitive revolution of the mid-1950s what Descartes was to the first phase of the age of modern philosophy.

Biography

Noam Chomsky was born in Philadelphia, Pennsylvania on December 7, 1928. His father, William Chomsky, was a noted Hebrew scholar who wrote *Hebrew, the eternal language*. While studying at the University of Pennsylvania under the supervision of Zellig Harris, he was also a Junior Fellow of the Harvard Society of Fellows from 1951 to 1955. After receiving a Ph.D. in linguistics from the University of Pennsylvania in 1955, Chomsky joined the faculty at MIT under the sponsorship of Morris Halle and was promoted to full professor of Foreign Languages and Linguistics in 1961, appointed Ferrari Ward Professor of Linguistics in 1966, and Institute Professor in 1976. Since 1967, when the University of London and the University of Chicago awarded him his first two honorary doctorates, he has been the recipient of scores of honorary degrees throughout the world. In 1969, he delivered the John Locke Lectures at Oxford University, and in 1970, the Bertrand Russell Memorial Lecture at Cambridge University. Among the almost countless other honors that he has received is the 1988 Kyoto Prize in basic science, created in 1984 for the purpose of recognizing outstanding achievements in categories not named by the Nobel Prizes. In 1984, he received the Distinguished Scientific Contribution Award of the American Psychological Association 'for enlarging our definition of scientific psychology'.

References

Anthony, Louise M., and Norbert Hornstein (eds.) 2003. Chomsky and his critics (Philosophers and their critics). Oxford: Blackwell Publishers.

Barsky, Robert F. 1997. Noam Chomsky: a life of dissent. Cambridge, MA: MIT Press.

Chomsky, Noam. 1957. Syntactic structures. The Hague: Mouton.

———. 1959. A review of B.F. Skinner's verbal behavior. Language 35. 26–58.

———. 1965. Aspects of the theory of syntax. Cambridge, MA: MIT Press.

———. 1966. Cartesian linguistics. Cambridge, MA: MIT Press.

———. 1972. Problems of knowledge and freedom: the Russell lectures. New York: Random House.

———. 1975. The logical structure of linguistic theory. New York: Plenum.

———. 1979. The morphophonemics of modern Hebrew. New York: Garland.

———. 1981. Lectures on government and binding. Dordrecht: Foris.

———. 1986. Knowledge of language. New York: Praeger.

———. 1995. The minimalist program. Cambridge, MA: MIT Press.

———. 2000. New horizons in the study of language and mind. Cambridge: Cambridge University Press.

———. 2000. Minimalist inquiries: the framework. Step by step: essays on minimalist syntax in honor of Howard Lasnik, ed. by Roger Martin, David Michaels, and Juan Uriagereka, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam, and Morris Halle. 1968. The sound pattern of English. New York: Harper and Row.

Harris, Zellig. 1951. Methods in structural linguistics. Chicago: University of Chicago Press.

Jenkins, Lyle. 2000. Biolinguistics: exploring the biology of language. Cambridge: Cambridge University Press.

Lyons, John. 1970. Chomsky, 3rd edition. London: Fontana.

McGilvray, James. 1999. Chomsky: language, mind, and politics. Cambridge: Polity Press.

Otero, C. P. (ed.) 1994. Noam Chomsky: critical assessments (8 tomes). London: Routledge.

Smith, Neil. 1999. Chomsky: ideas and ideals. Cambridge: Cambridge University Press.

HOWARD LASNIK AND CARLOS OTERO

See also **Bar-Hillel, Yehoshua; Computational Linguistics; Grammar, Theories; Structuralism**

Clark, Eve V.

Educated in France and the United Kingdom, Eve V. Clark started her linguistics career in the United States at Stanford University. She worked on the Language Universals Project led by Joseph Greenberg and Charles Ferguson and became one of the first faculty members of the Linguistics Department. She has worked extensively in various subfields of linguistics including pragmatics, lexical semantics, and psycholinguistics. Her greatest contribution to linguistics and psychology is her groundbreaking research on meaning acquisition—in particular, her earlier work on the semantic feature hypothesis, her comparative studies of children's word formation, and her current research on adults' offers of and children's uptake of lexical and semantic information.

In the 1970s, Eve Clark's research focused mostly on the semantic feature hypothesis. Simply put, the semantic feature hypothesis states that in learning a new word, children only learn some of its semantic features and then add to this knowledge as they find out more about what the word means. In one of her earliest studies, Clark looked at the acquisition of the meaning of *before* and *after*. The acquisition order of these two words is as follows: (1) children did not understand either word; (2) they understood *before* but not *after*; (3) they interpreted *after* as if it meant *before*; and finally, (4) they understood both words correctly. Notice that the meaning of these two words is made up of two components—namely, Time and Prior, and these components can have either a positive

or negative value. A positive value of Time shows that the word refers to some aspect of time. Similarly, the values of Prior indicate that one event precedes (positive) and follows (negative) the other. In light of the semantic analysis of these two words, one can see that children do not understand the component Time at first. Once they understand that both *before* and *after* contain the component Time, they may still not have acquired the feature Prior. At the third stage, they understand the positive member of Prior (i.e. *before*) but not the negative member (i.e. *after*)—in other words, the positive value of Prior is learned before the negative one. Finally, they understand both. Clark showed that this type of analysis could explain the findings of other studies on the acquisition of other relational terms such as *more–less, high–low,* and *tall–short.*

In the 1980s, Clark's research focused on word formation in acquisition. With colleagues Ruth Berman, Barbara Hecht, and Randa Mulford, she conducted extensive comparative studies of children's word formation in English, French, and Hebrew. She proposed several principles—semantic transparency, productivity, and conventionality—that influence children's acquisition of word-formation devices (e.g. compounding, affixation) in language. For instance, younger English-speaking children (e.g. aged three) often rely on simple compounds (e.g. *plant-man*) to coin new nouns; their use of *-er* for agents (e.g. *wall-builder*) is rather inconsistent.

This finding is in accordance with the principle of semantic transparency: while the head noun *-man* is semantically transparent to them (i.e. they know what the noun *man* means), younger children still need to analyze *-er* to obtain the agentive meaning of the suffix. As children get older, they replace words that they have coined through compounding with those conventionally used (e.g. *plant-man* with *gardener*). This provides evidence for the principle of conventionality: 'For certain meanings, a conventional word or word formation device exists that should be used in the language community' (Clark and Berman 1984:549). Another principle is productivity: [T]hose word-formation devices used most often by adults in word innovations are preferred in the language for constructing new word forms' (ibid.: 548). Empirical evidence shows that children rely on this in acquiring word-formation devices. For instance, among the agentive suffixes *-er*, *-ist*, and *-ian* in English, *-er* is the most frequent. According to the principle of productivity, children should acquire *-er* before the others, and this turns out to be the case (see Clark and Cohen 1984). Cross-linguistic studies have shown that regardless of the language being acquired, these principles play a significant role in children's acquisition and use of word formation.

Since the 1990s, Clark's research has centered on the pragmatics of child-directed speech, especially its relation to children's meaning acquisition. In acquiring vocabulary, children need to learn to link specific terms with their conventional meanings. In the last few decades, some researchers have proposed that children observe certain built-in constraints in mapping meanings onto word forms, e.g. the mutual exclusivity constraint where each referent is picked out by just one word. However, studies have shown that the constraint approach underestimates the importance of adult contribution. Children as young as two years old can and do make use of lexical and semantic information offered to them by adults. For example, two-year-olds readily accept and use multiple terms for the same referent (e.g. *dog* and *spaniel*) when adults provide them with this information. An alternative to the constraint approach is to examine the role of adult-directed speech in meaning acquisition and children's uptake of adults' offers of words and relations among words. Clark's current research goals are to (1) establish the range of adults' offers of words and relations and (2) determine how effective different types of offers are for word learning, as measured by children's uptake.

The central themes in much of Clark's research are the emphases on general cognitive principles and on social interaction in language acquisition. While many studies on meaning acquisition focus on specific constraints, the principles that Clark proposed for meaning acquisition apply to other facets of language acquisition and to adult language use as well. For instance, the pragmatic principles of conventionality and contrast (i.e., a difference in meaning is marked by a difference in form) are not only essential to children's acquisition of meaning but also guide adult speakers in their uses of language. The explanatory power of these principles reflects the strength of Clark's research.

Biography

Eve V. Clark was born in Camberley, UK, July 26, 1942. She received her M.A. Hons. (1965) in French Language and Literature, with minors in Spanish and Phonetics, University of Edinburgh; Postgraduate Diploma in General Linguistics (1966), University of Edinburgh; and Ph.D. in Linguistics (1969), University of Edinburgh. She was Assistant Instructor and Instructor in the Department of French at the University of Pittsburgh (1967–1969); Research Associate, Language Universals Project, Stanford University (1969–1970); Lecturer, Assistant Professor, Associate Professor, and Full Professor, Stanford University (1970–present); Visiting Scientist, Max-Planck-Institut für Psycholinguistik Nijmegen, The Netherlands (1981, 1983–1984, 1990–1991, 1997–1998); Fellow, Center for Advanced Study in the Behavioral Sciences, Stanford, California (1979–1980); Guggenheim Fellow, John Simon Guggenheim Memorial Foundation (1983–1984); and Foreign Member, Royal Netherlands Academy of Arts and Sciences (KNAW) (elected 1991).

References

Clark, Eve. 1971. On the acquisition of the meaning of before and after. Journal of Verbal Learning and Verbal Behavior 10. 266–75.

———. 1973. What's in a word? On the child's acquisition of semantics in his first language. Cognitive Development and the Acquisition of Language, ed. by Timothy E. Moore, 65–110. New York: Academic Press.

———. 1983. Meaning and concepts. Handbook of child psychology, Vol. 3: Cognitive development, 4th edition, ed. by John H. Flavell and Ellen M. Markman, 787–840. New York: Wiley.

———. 1991. Acquisitional principles in lexical development. Perspectives on thought and language: interrelations in development, ed. by Susan A. Gelman and James P. Byrnes, 31–71. Cambridge: Cambridge University Press.

———. 1993. The lexicon in acquisition, Cambridge and New York: Cambridge University Press.

———. 1999. Acquisition in the course of conversation. Studies in the Linguistic Sciences 29. 1–18.

Clark, Eve, and Ruth Berman. 1984. Structure and use in the acquisition of word-formation. Language 60. 547–90.

Clark, Eve, and Sophia Cohen. 1984. Productivity and memory for newly formed words. Journal of Child Language 11. 611–25.

Donaldson, M., and G. Balfour. 1968. Less is more: a study of language comprehension in children. British Journal of Psychology, 59. 461–471.

Donaldson, M., and R. Wales. 1970. On the acquisition of some relational terms. Cognition and the development of language, ed. by J.R. Hayes, 235–268. New York: Wiley.

Gelman, S., J. Coley, K. Rosengren, E. Hartman, and A. Pappas. 1998. Beyond labeling: the role of maternal input in the acquisition of richly structured categories. Monographs of the Society for Research in Child Development. 63.

Markman, E. 1987. How children constrain the possible meanings of words. Concepts and conceptual development: ecnological and intellectual factors in categorization, ed. by U. Neisser, 255–287. Cambridge: Cambridge University Press.

Markman, E., and G. Wachtel. 1988. Children's use of mutual exclusivity to constrain the meanings of words. Cognitive Psychology 20. 121–157.

Merriman, W., and L. Bowman. 1989. The mutual exclusivity bias in children's word learning. Monographs of the Society for Research in Child Development. 54.

ANDREW WONG

See also **Acquisition; Pragmatics; Psycholinguistics; Semantics**

Clause

Sentences can be of arbitrary length. Every well-formed sentence, however, can be adequately described in terms of its internal structure: the words it contains, the phrases these build and the morphemes they are made up from, and other such relations. Beyond the regular phrase level—noun phrase (NP), verb phrase (VP), etc.—sentences can be divided into clauses. Above all, a sentence is always a clause; but some sentences may consist of more than one clause.

In traditional grammar, the distinction is made between *main* (subordinate) and *subordinate* (dependent) clauses. In a sentence like *John likes Mary*, the whole structure is also one clause, but in the slightly longer sentence *John likes Mary **when she plays football,*** the boldfaced part is a separate clause all by itself. In generative grammar, the term *embedded* is used to indicate that the subordinate clause is embedded under the main clause, but still part of the same sentence. This is more than mere notation: traditional grammar does not consider subordinate clauses to be clauses proper; in the generative approach, however, all well-formed sentences are clauses, and every (grammatical) embedded clause is by definition a well-formed structure on the sentential level, hence a clause.

The sentence–clause distinction in generative terms (as in the Principles-and-Parameters theory) is often one of categorial projection, selection, and complementation. Two main categories are decisive for the type of structure: the Inflectional Phrase (or IP for short, formerly S) and the Complementizer Phrase (CP or S'). IP denotes the traditional concept that a well-formed sentence consists of a subject and a predicate (often the verb phrase VP). In X'-theory, the canonical subject position is the specifier of IP. The sentence *John likes Mary* can adequately be described in terms of an IP (and further internal structure). The CP-level is invoked for additional material, such as the question *Who does John like?*, which is transformationally derived from the former sentence. Here, the subject *John* occupies the same position in a structural description (i.e. the tree or the phrase-marker) of the sentence as the original sentence. But it is assumed that the object *who* has moved from its canonical predicate- or VP-internal position to the beginning of the sentence, the specifier of CP.

We can thus see that IP and CP can be well-formed main clauses. Finiteness plays a role for, but is no essential identifier of, main vs. subordinate clauses. In principle, any clause can be finite or nonfinite. As a rule of thumb, unmarked, declarative sentences—what have sometimes been dubbed *kernel sentences*—are IPs, while more complicated or derived structures (including interrogatives, imperatives, or exclamatives, such as *Does John likes Mary?*, *Play better soccer now!,* or *What a great game she played!*) are usually CPs. Another rule of thumb is that a well-formed clause contains only one subject and one predicate (i.e. one main verb and its potential complements, depending on whether the verb is an intransitive, a transitive, or a ditransitive verb, for example). But virtually every clause can be embedded under or subordinated to a higher clause. This depends on the type of clause and/or on the potential main clause predicate.

Addressing the former first, subordinate or embedded clauses come in various types. We commonly

distinguish between *adverbial*, *complement,* and *relative* clauses. An adverbial clause (also *oblique clause*) relates to the main clause through the semantic meanings expressed by adverbs, such as time, manner, place, instrument, circumstance, concession, purpose, result, cause, or condition. In English, adverbial clauses are typically introduced by a *subordinating conjunction* (or *subordinator*), an element that links the adverbial clause to the higher main clause. An example is *John likes Mary because she plays soccer well*. Other subordinators are *after*, *when*, *whenever*, *while*, *as*, *although*, or *if*. In generative analysis, subordinators invariably head a CP-projection; adverbial clauses are thus CPs.

Complement clauses are characterized as clauses that serve as a complement to a lexical item, i.e. without which that lexical item would not be complete. This addresses the role of the potential main clause predicate mentioned above. Two major classifications of complement clauses exist, namely noun-complement and verb-complement clauses. Just as some physical object must be expressed to form a complete structure of the verb *throw* (as in *The goalkeeper threw **the ball***), as part of the definition or subcategorization frame of that verb, some clause must be expressed to satisfy the subcategorization frame of a verb like *tell*. The boldfaced part in *Mary told John **(that) she played soccer*** is a complement clause, namely complement to the main verb *tell*. (Naturally, *tell* can also take an NP as a complement, as in *Mary told John **a lie***, but we concentrate on clauses here.) In this case, the complementizer *that* is optional; it can be sued or left out. In this instance, other elements could also be used instead, such as *why* or *how*. The important relation is that between the main verb *tell* and the proposition expressed by the following clause. A similar relationship can be found with certain nouns. A noun like *fact* takes a complement clause. For example, we say *John likes the fact **that Mary plays soccer***. It is thus part of the subcategorization frame of a noun like *fact* that it must be followed by a complement clause.

Relative clauses, finally, modify, describe, or further specify a noun, but are structurally part of the entire noun phrase, i.e. the *head noun*, any other modifiers of it, and the relative clause form one constituent. Moreover, relative clauses are optional, not obligatory. Thus, while in a noun-complement clause, the complement clause must be used, relative clauses need not be. Consider, for example, the sentence *John likes the girl **who plays soccer***. The boldfaced part is a relative clause, adding further information about the NP, *the girl*. This type is a *restrictive* relative clause, which is characterized by being essential for identification (otherwise we would not know which girl John likes). A *nonrestrictive* relative clause is not required for identification of the NP, as in *Mary, **who plays soccer**, is a nice girl*.

To sum up, the property of human language that sentences can be of infinite length does not mean that longer sentences cannot be analyzed. Rather, a property of language is that sentences can consist of more than one clause, but each clause can be clearly identified. We distinguish between a (unique) main clause and (any number of) subordinate clauses. The latter come in three types: adverbial, complement, and relative clauses.

References

Haegeman, Liliane. 1991. Introduction to government and binding, 1st edition. Oxford: Blackwell; 2nd edition, 1994.

Huddleston, Rodney. 1984. Introduction to the grammar of English. Cambridge: Cambridge University Press.

Huddleston, Rodney, and Geoffrey K. Pullum. 2002. The Cambridge grammar of the English language. Cambridge: Cambridge University Press.

Keenan, Edward. 1985. Relative clauses Vol. 2: Complex constructions, ed. by Shopen, pp. 141–70.

Noonan, Michael. 1985. Complemenatation. Language typology and syntactic description, Vol. 2: Complex constructions, ed. by Timothy Shopen, 42–140. Cambridge: Cambridge University Press.

Radford, Andrew. 1988. Transformational grammar. Cambridge: Cambridge University Press.

Shopen, Timothy (ed.) 1985. Language typology and syntactic description, Vol.2: Complex constructions. Cambridge: Cambridge University Press.

Thompson, Sandra A., and Ronald E. Longacre. 1985. Adverbial clauses. ed. by Shopen, pp. 171–234.

KLEANTHES K. GROHMANN

See also **Grammar, Traditional**

Clause-Type Indicators

Communicating by language is a highly structured and goal-directed social activity that serves many purposes: making contact, delivering a speech, telling a joke, expressing surprise, gathering information, giving orders, and so forth. For verbal interaction to be effective, the speaker has to formulate what he or she wants

to say in such a way that its communicative intention is recognizable or identifiable to the other participants in the speech situation. Making a statement will naturally take a different form than posing a question or issuing a command. In language theory, the systematic relation between the grammatical form of sentences and their conventionalized conversational use is called clause type.

The three major clause types that occur most frequently in the world's languages are declaratives, interrogatives, and imperatives, which correlate with statements of facts, inquiries, and directives (a cover term for commands, requests, orders, and the like), respectively. Clause types represent a grammatical system in at least two respects. First, one can easily construct triplets of corresponding declaratives, yes–no questions, and imperatives (e.g. *John left in a hurry; did John leave in a hurry?; leave us alone!*). Second, clause types are mutually exclusive, no sentence belonging simultaneously to two different types.

To distinguish one clause type from another on a formal basis, languages resort to a large variety of grammatical devices, which include changes in the shape of verbs, word-order alternations, different intonation patterns, and special particles and affixes. The concern here is with such dedicated affixes and particles, which are put to use as clause-type indicators. Clause-type indicators constitute a subset of function words with a primarily classificatory function that mark the sentence they modify as a declarative or an interrogative construction. Out of the ordinary as sentence-type indicators may look at first sight, they are also attested in English. Consider, for instance, the subordinating conjunctions *that* and *whether*, which introduce embedded declarative and interrogative sentences, respectively (e.g. *John said **that** there would be enough food at the party* vs. *John wondered **whether** there would be enough food at the party*).

As many linguists have observed, declarative sentences tend to be the most unmarked of these clause types: they typically occur without any clause type indicator or specific ordering and impose the fewest restrictions on what verbal categories can be selected. There are, however, many languages with marked declaratives. For example, Welsh (a Celtic language of the British Isles) has a clause-initial particle *y(r)*, which is placed in declarative constructions with periphrastic verb forms, which are composed of the present and the imperfect forms of *bod* 'to be' (third-person masc. sing. *mae* 'he is') and a verbal noun (e.g. *Y mae Siôn yn gweld draig* [PARTICLE is John PROGRESSIVE.PARTICLE see _{VERBAL-NOUN}] 'John sees (lit. is seeing) a dragon'). The declarative particle *y(r)* contrasts with other sentence-type indicators like the negative particle *ní(d)* and the question particle

a, neither of which can occur together with *y(r)*. Yet, the particle *y(r)* is absent in embedded and verb-initial declarative constructions. In Maale (an Omotic language of Southern Ethiopia), all clause types are marked by a special verbal affix. Thus, the main verb of affirmative declarative clauses contains the affix *–ne* (e.g. *ʔ–atsí ziginó mukk–é–**ne*** [man yesterday come-PERFECTIVE-AFFIRMATIVE. DECLARATIVE] 'The man came yesterday'), while the corresponding interrogatives are characterized by the affix *–iya* when the main verb is marked for Perfective aspect, which indicates the completion of the event that is described (e.g. *ʔ–atsí ziginó mukk–é–**iya***? [man yesterday come-PERFECTIVE-INTERROGATIVE] 'Did the man come yesterday?').

A more complex situation is obtained in languages with so-called evidential systems. Evidential particles (or affixes) indicate the source reliability of the speaker's knowledge and the kind of evidence he or she adduces for what is being communicated. In languages with elaborate evidential marking, such as Tuyuca (a Tucanoan language spoken in Colombia and Brazil), there are no unmarked declaratives, since every sentence must contain an evidential marker qualifying the information on which an assertion is based. A simple declarative like *he played soccer* corresponds to five different verbal constructions, each associated with a specific evidential marker. Compare: *díiga apé–**wi*** [soccer play-VISUAL] 'he played soccer' (I say him play), *díiga apé–**ti*** [soccer play-NON-VISUAL] 'he played soccer' (I heard the game and him, but I didn't see it or him), *díiga apé–**yi*** [soccer play-APPARENT] 'he played soccer' (I have evidence that he played soccer, but I didn't actually see the game), *díiga apé–**yigɨ*** [soccer play-HEARSAY] 'he played soccer' (I obtained the information from someone else), *díiga apé–**hîyi*** [soccer play-DEDUCTIVE] 'he played soccer' (it is reasonable to assume it).

Interrogative sentences come in two varieties: yes–no questions, where the truth of the questioned statement is at issue (e.g. *Was John invited to the party?*) and the constituent question, where the question word or phrase signals the missing piece of information (e.g. ***Who was invited to the Party?***).

In Polish, the question marker *czy* is used to distinguish main and embedded yes-or-no questions from declarative sentences (e.g. *Pan Kowalski był w Austrii* 'Mr. Kowalski has been to Austria' vs. ***czy** Pan Kowalski był w Austrii* 'Has Mr. Kowalski been to Austria?' and *nie pytałem **czy** Pan Kowalski był w Austrii* 'I did not ask <u>whether</u> Mr. Kowalski had been to Austria'). In Mandarin Chinese, question words like *shei* and *shemne* convey not only a question interpretation 'who' and 'what' but may also be interpreted as indefinite pronouns 'someone, anyone' and 'something, anything' (e.g. ***shei** mai–le yi–ben-shu*? [who

buy-ASPECT one-CLASSIFIER -book] '<u>who</u> bought a
book?' vs. *ta bu xiang jian **shei*** [S/he not want see
who] 'She (or he) does not want to see anyone'). If, on
the other hand, the question particle *ne* occurs in sen-
tence-final position, only the interrogative and not the
indefinite interpretation of the question word is avail-
able (e.g. Hufei *mai–le shemne ne*? [Hufei buy-ASPECT
what INTERROGATIVE.PARTICLE] '<u>What</u> did Hufei buy?'
[NOT 'Hufei bought <u>something</u>']). Languages with
dedicated interrogative particles like Chinese leave the
question word in exactly the same syntactic position as
the corresponding declarative sentence, while lan-
guages without such clause-typing devices, like
English, place the question word in front of the clause
and require the inversion of the auxiliary verb and sub-
ject (and insertion of *do* if there is no auxiliary verb).

With evidential particles and affixes being a relative-
ly recent discovery, more research needs to be done to
clarify the functional role of such clause-type indicators.

References

Amha, Azeb. 2001. The Maale language. Ph.D. Dissertation. Leiden University, CNWS Publications 99.

Barnes, Janet. 1984. Evidentials in the Tuyuca verb. International Journal of American Linguistics 50. 255–71.
Bielec, Dana. 1998. Polish—an essential grammar. London and New York: Routledge.
Cheng, Lisa Lai-Shen. 1991. On the typology of wh-ques-tions. Ph.D. Dissertation, Massachusetts Institute of Technology. (reprinted in 1997, New York and London: Garland).
Cinque, Guglielmo. 1999. Adverbs and functional projections: a cross-linguistic perspective. Oxford and New York: Oxford University Press.
Palmer, Frank R. 1986. Mood and modality. Cambridge and New York, Cambridge University Press.
Sadock, Jerrold M., and Arnold M. Zwicky. 1985. Speech act distinctions in syntax. Language typology and syntactic description, Vol. I: Clause structure, (ed. by Timothy Shopen,) 155–96. Cambridge and New York: Cambridge University Press.
Sproat, Richard. 1985. Welsh syntax and VSO structure. Natural Language and Linguistic Theory 3. 173–216.
Willett, Thomas. 1988. A cross-linguistic survey of the gram-maticalization of evidentiality. Studies in Language 12. 51–97.
Yule, George. 1996. Pragmatics, Oxford introductions to lan-guage study. Oxford and New York: Oxford University Press.

CHRIS H. REINTGES

See also **Clause**

Clinical Linguistics

Clinical linguistics, a branch of applied linguistics, is
the use of linguistics to describe, analyze, and treat
language disability. The study of linguistic aspects of
communication disorders is of relevance to a broader
understanding of language and linguistic theory.
Practitioners of clinical linguistics range from aca-
demic linguists with research and teaching interests in
language disability to practicing professionals such as
speech and language pathologists/therapists, educa-
tional and clinical psychologists, and neurologists.
Research in the field tends to be multifaceted, drawing
on a wide range of disciplines in addition to linguis-
tics, such as psycholinguistics, cognitive science, neu-
roscience, and biomedical science.

Although systematic phonetic descriptions of speech
disorders have been routine since the 1950s, clinical lin-
guistics did not emerge as a coherent discipline inclu-
sive of phonology, grammar, semantics, and pragmatics
until the late 1970s. This was in part due to the pioneer-
ing work of David Crystal whose book *Clinical linguis-
tics* (1981) has been particularly influential in defining

the area. Other milestones were the establishment of the
journal *Clinical linguistics and phonetics* in 1987 and
the founding of the International Clinical Phonetics and
Linguistics Association (ICPLA) in 1991.

Rather than diagnosing using medical criteria, clini-
cal linguistics focuses on the linguistic manifestations
of the disorder, aiming to provide a comprehensive
typology of language disorders based on their linguistic
characteristics. This approach aids in diagnosing and
treating impairments of unknown underlying causes.
For example, there is as yet no agreed medical explana-
tion for Specific Language Impairment (SLI), a condi-
tion found in children who have problems with spoken
language but no other obvious cognitive or neurological
deficit. In spite of this, it is still possible to describe the
linguistic characteristics of SLI precisely enough for
research purposes and for devising remedial programs.

There appears to be no level of language organiza-
tion that is immune to impairment. Language disor-
ders are equally likely to be found in adults and in
children who are still acquiring language, and in both

213

the production and comprehension of spoken, written, and signed language. Anatomical deficits such as cleft palate may affect a speaker's phonetics and phonology, resulting in an inability to differentiate pairs of words such as 'bat' and 'mat'. Grammar, semantics, and pragmatics may be impaired as a result of developmental disorders in children or a stroke in adults. A child with SLI might repeat 'Has the mouse been chased by the cat?' as 'A mouse chasing cat'. A stroke patient might see nothing wrong with adding a past tense suffix to nouns, resulting in words like 'towned' and 'faithed', due to problems with word formation. People with autism find it difficult to make use of context to infer nonliteral meaning.

Any linguistic theory that purports to throw light on the nature of the human mind should describe both the normal and the pathological. Some have argued that certain syntactic anomalies found in the language of people with Broca's aphasia, a condition resulting from damage to Broca's area of the brain, can be explained using Chomsky's 'Principles and Parameters' theory of syntactic structure. These researchers believe that such anomalies provide evidence that syntactic abilities are separate from other cognitive skills. In addition, data from children with SLI and Williams Syndrome, a genetic condition in which linguistic proficiency develops despite poor cognitive abilities, have been use d to support a similar view of syntax as an autonomous mental 'module'. Others, however, cite evidence from a range of language disorders in support of a nonmodularist, functionalist account of language suggesting a considerable degree of codependency between linguistic and cognitive processes. Clinical linguistic research will continue to play a key role in debates of this kind.

The range of analytical methods used in clinical linguistics is comparable to that found in other areas of linguistics. Both theory-based and data-driven approaches are used, with a wide selection of materials produced for purposes of diagnosis, assessment, and intervention. Nevertheless, clinical linguistics is still a new field and a great deal of exploratory work remains to be done, particularly in the form of case studies and in areas such as semantics. Often, it is only when a complex system goes wrong that we become aware of the contribution—and even the existence—of its component subsystems. One potential growth area for clinical linguistics, therefore, is its role in informing and evaluating linguistic theory and in illuminating our understanding of normal language structure and use.

References

Ball, Martin J. 1989. Phonetics for speech pathology. London: Taylor & Francis, 2nd edition, London: Whurr, 1993.

Ball, Martin J., and Raymond D. Kent (eds.) 1997. The new phonologies: developments in clinical linguistics. San Diego: Singular.

Chiat, Shula. 2000. Understanding children with language problems. Cambridge and New York: Cambridge University Press.

Crystal, David. 1980. Introduction to language pathology. London: Edward Arnold, 4th edition (co-author Rosemary Varley), London: Whurr, and San Diego: Singular, 1998.

Crystal, David. 1981. Clinical linguistics. Vienna: Springer-Verlag, London: Whurr, 1989.

Grodzinsky, Yosef. 1990. Theoretical perspectives on language deficits. Cambridge, MA: MIT Press.

Grundy, Kim (ed.) 1989. Linguistics in clinical practice. London: Taylor & Francis; 2nd edition, London: Whurr and San Diego: Singular, 1995.

Hewlett, Nigel, Louise Kelly, and Fay Windsor (eds.) 2001. Themes in clinical linguistics and phonetics. New York: Lawrence Erlbaum.

Howard, Sara J., and Barry C. Heselwood. 2005. Clinical phonetics and phonology: analysing speech in a clinical context. New York: Academic Press.

Perkins, Michael R. 2005. Pragmatics and communication impairment. Cambridge and New York: Cambridge University Press.

Perkins, Michael R., and Sara J. Howard (eds.) 1995. Case studies in clinical linguistics. London: Whurr and San Diego: Singular.

MICHAEL PERKINS

See also **Aphasia; Applied Linguistics: Overview; Psycholinguistics**

Code-Switching

Speakers of more than one language (e.g. bilinguals) are known for their ability to code-switch or mix their languages during communication. This phenomenon occurs when bilinguals substitute a word or phrase from one language with a phrase or word from another language. While some linguists suggest that people code-switch as a strategy in order to be better understood and to enhance the listeners' comprehension, code-switching among

bilinguals has traditionally been viewed as a strategy to compensate for diminished language proficiency. The premise behind this theory is that bilinguals code-switch because they do not know either language completely. This argument is also known as semilingualism, meaning bilinguals 'almost' speak both languages correctly. However, language proficiency is not clearly defined, as it is unclear whether reading and writing language skills should take precedence over spoken skills in determining language proficiency. This reliance on reading and writing is problematic because most bilinguals receive their formal education in one language, while a majority of their social interactions take place in the other language. Thus, when their reading and writing abilities are tested in both languages, the language in which bilinguals received more formal education will usually fare better.

Recent psycholinguistic research has focused on how code-switching is a natural product of the interaction of the bilingual's two languages. Early researchers viewed code-switching as evidence that the bilinguals' two languages were organized in separate and distinct mental dictionaries. For example, a general finding throughout the literature is that bilinguals take longer to read and comprehend sentences containing code-switched words as compared to monolingual sentences. Apparently, this time-consuming process is due to a 'mental switch mechanism' that determines which of the bilingual's two mental dictionaries are 'on' or 'off' during the course of language comprehension. This mental switch is responsible for selecting the appropriate mental dictionary to be used during the comprehension of a sentence. Other research shows that bilinguals comprehend code-switched words faster when there is an overlap between the two languages' sound systems. For example, Chinese–English bilinguals, where Chinese is the native language, take longer to recognize English code-switched words in Chinese sentences, but only if the English words begin with consonant–consonant clusters (e.g. block), as opposed to consonant–vowel clusters (e.g. big), because the Chinese language words do not begin with consonant–consonant clusters. Other important factors reported to influence the recognition of code-switched words include context, phonetics, homophonic (e.g. words pronounced the same) and homographic (e.g. words spelled the same) overlap between the two languages.

Another current view suggests that language dominance, which language is used more frequently, plays an important role in code-switching. For example, Spanish–English bilinguals report more linguistic interference (code-switching) when they communicate in Spanish, their first language, and little or no code-switching when they communicate in English, their second language. In other words, these bilinguals code-switch more when they communicate in Spanish than when they use English. Psycholinguistic evidence also suggests that bilinguals retrieve English code-switched words faster when they listen to Spanish sentences, whereas they are slower to retrieve Spanish code-switched words as they listen to English sentences. This evidence suggests that the bilingual relies on the second language as opposed to the first. How are these findings explained? The general premise behind this view is that after a certain level of fluency and frequent use of the second language, a language shift occurs. That is, during early stages of bilingualism, Spanish–English bilinguals rely on their first language when they communicate in their second language. As a result, bilinguals are more likely to code-switch to Spanish, when they communicate in English. However, as the second language becomes the dominant language, bilinguals rely on the second language when they communicate in the first language. In this case, bilinguals code-switch to English when they communicate in Spanish. The second language becomes more readily accessible and bilinguals come to rely on it more. Regardless of which language the bilingual learned first, the more active (dominant) language determines which mental dictionary is going to be accessed faster. This argument is reasonable since most bilinguals in the United States, whose first language is Spanish, obtain their formal education in English. Likewise, many of their everyday interactions involve the second language. As a result, words and concepts in English, the second language, become more accessible than words in Spanish, the first language. Thus, code-switching is not the same for both languages. Rather, it depends on language dominance.

In short, code-switching among bilinguals may be indicative of difficulties in retrieval (access) affected by a combination of closely related factors such as language use (i.e. how often the first language is used) and word frequency (i.e. how much a particular word is used in the language). Examination of code-switching behavior can contribute to a further understanding of second-language acquisition, as well as language acquisition and development, and general linguistic theory.

References

Lederberg, Amy R., and Cesareo Morales. 1985. Code switching by bilinguals: evidence against a third grammar. Journal of Psycholinguistic Research 14. 113–36.

Li, Ping. 1996. Spoken word recognition of code-switched words by Chinese–English Bilinguals. Journal of Memory and Language 35. 757–74.

Macnamara, John, and Seymour Kushnir. 1971. Linguistic independence of bilinguals: the input switch. Journal of Verbal Learning and Verbal Behavior 10. 480–87.

Heredia, Roberto R. 1997. Bilingual memory and hierarchical models: a case for language dominance. Current Directions in Psychological Science 6. 34–9.

Kolers, Paul. 1966. Reading and talking bilingually. American Journal of Psychology 3. 357–76.

Myers-Scotton, Carol. 1993. Duelling languages: grammatical structure in code-switching. New York: Oxford University Press.

ROBERTO R. HEREDIA AND JEFFREY M. BROWN

See also **Acquisition; Bilingual Acquisition; Bilingualism**

Coherence in Discourse

Discourse is a communicative event in which language plays a prominent role. It minimally requires a sender (writer, speaker), a receiver (reader, listener), and a message that is being communicated. This message is not just a concatenation of clauses; it forms a unified, coherent whole. Both the sender and receiver normally have the implicit agreement that the message being communicated is coherent.

Coherence in discourse has been studied in a range of disciplines, including linguistics, philology, sociology, philosophy, psychology, and computer science. Linguists identify and analyze inventories of the linguistic markers of coherence that are available in a language. Sociologists explore the production and comprehension of coherent discourse in naturalistic conversations that involve different groups and cultures. Psychologists collect data in experiments that test hypotheses about the effect of coherence on cognitive processing and representations. Computer scientists design and test computer models that attempt to produce and test coherent text.

The term *coherence* has been defined in various ways. Some researchers apply the term *cohesion* to the surface structure of the text and the term *coherence* to the concepts and relations underlying its meaning. *Cohesion* has sometimes been applied to smaller units of language in the text, and *coherence,* to some general overall interrelatedness in the text. Other researchers have defined *cohesion* as continuity in word and sentence structure, and *coherence* as continuity in meaning and context. As in the case of *coherence*, *discourse* has been defined in different ways. Several years ago, the term *discourse* was reserved for dialogue, and *text* was reserved for monologue. In contemporary research, *discourse* covers both monologic and dialogic spoken and written language.

Somewhat more subtle distinctions are sometimes made. One can distinguish between discourse-as-product (the linguistic construct) and discourse-as-process (the communicative event). *Coherence* can be reserved for the conceptual relationships that comprehenders use to construct a coherent mental representation accommodated by what is said in the discourse. *Cohesion* is limited to the linguistic markers that cue the comprehender on how to build such coherent representations. Cohesion emphasizes discourse-as-product, and coherence emphasizes discourse-as-process.

Cohesion alone is not sufficient for the interpretation of the discourse. Comprehenders generate inferences on the basis of background knowledge and discourse constraints. Much of the background knowledge is experiential; hence, it involves common procedures and activities (called *scripts*), social interactions, and spatial settings. For instance, a narrative usually describes a setting, an action sequence with a conflict and plot, and an outcome. A script for eating in a restaurant would furnish inferences and help coherently tie together the explicit content of a narrative about a bad restaurant experience. Although cohesion alone cannot fully account for coherence in discourse, the psycholinguistic literature has shown that cohesion facilitates coherence.

Cohesion and coherence can be divided into local (microstructure) and global (macrostructure). Local cohesion and coherence are related to the interrelatedness between adjacent discourse segments. Global cohesion and coherence are related to the interrelatedness of larger spans of discourse. For instance, scripted action sequences are globally coherent. Also, there are the rhetorical structures of narrative (such as setting + conflict + plot + resolution), expository (such as claim + evidence, problem + solution), and other discourse genres.

Cohesion and coherence can be grammar driven and vocabulary driven. Grammar-driven cohesion refers to sentence structure, word structure, and the intonation of the discourse segments. Vocabulary-driven cohesion

refers to the lexical vocabulary of the discourse segment. These cohesion cues activate vocabulary-driven (pre-grammatical, knowledge-based) and grammar-driven (syntax-based) coherence. Vocabulary-driven and grammar-driven coherence are not necessarily mutually exclusive but often support each other, as illustrated below.

Consider the sentence **The dean** (i) *read* **the New York Times** (ii) *in* **his office** (iii). A paraphrase with grammar-driven cohesion would reduce the discourse elements to the grammatical necessities: **He** (i) *always reads* **it** (ii) **there** (iii). A vocabulary-driven paraphrase, on the other hand, would find meaningful lexical alternatives, as in **The man** (i) *always reads* **the newspaper** (ii) **behind his desk** (iii).

In addition to the distinctions between local and global and between grammar- and vocabulary-driven cohesion, the types of cohesion discussed below have often been recognized.

Conjunctions relate adjacent discourse segments. There have been several classifications of these conjunctions in virtually every field. Most of these classifications include additive (*and*, *but*), temporal (*before*, *until*), and causal (*because*, *although*) conjunctions that are either extensive (*and*, *before*, *because*) or adversative (*but*, *until*, *although*).

Coreference specifies that two expressions refer to the same entity. Often, the coreference is grammar driven by the use of pronouns, both pronominal (*he, she*) and reflexive (*himself, herself*). The interpretation of these pronouns is determined by their antecedents, i.e. previously mentioned words referring to the same person or object. The coreference can be both forward and backward. *Anaphoric* reference is a backward reference to an antecedent noun phrase or clause that was introduced earlier in the discourse (**John** *kissed* **Mary** *because* **he** *loved* **her**). *Cataphoric* reference is a forward reference to a noun phrase or clause that will be mentioned later in the text (*Because* **he** *loved* **her**, **John** *kissed* **Mary**).

With substitution, repeated forms, and ellipsis, a constituent of one expression is replaced by a constituent of another (substitution), is repeated (repeated forms), or is omitted (ellipsis). The intended meanings can be reconstructed from the preceding discourse and from world knowledge.

Will we make it on time?

(1) *I think so* (substitution of *we will make it in time* by *so*).
(2) *Yes, we will make it on time* (repeated forms).
(3) *If we hurry* (ellipsis: *we will make it on time* is omitted).

With lexical relationships, the type of cohesion is vocabulary driven. Two lexical items are related to each other insofar as they mean the same thing (synonyms) or the opposite thing (antonyms), stand in a superset/subset relationship (hypernym vs. hyponym, respectively), or have some other conceptual relationship.

(1) **The tax collector** *sent another letter*.
(2) *I don't like* **this guy**.
(3) **That monster** *never leaves us alone.*
(4) **The sweetheart** *keeps asking for more each year*

With comparison, a constituent in an expression is compared with a constituent in another expression (*I like the oak cabinet. The pine desk is much nicer*).

Discourse psychologists have extensively investigated five cohesion and coherence relations that are related to the previously mentioned seven: referential, spatial, causal, temporal, and additive relationships. They answer the questions of the who, where, why, when, and what of the events described by the discourse. Explicit markers facilitate the comprehension process.

Several classifications of relations have been proposed. Some focus only on the closed set of grammar-driven cohesion, whereas others include vocabulary-driven relations and relations that are reconstructed from world knowledge and the unique situation conveyed in the text. Those classifications that go beyond grammar consider the intentions of the producer of the communicative event. In written monologic discourse, comprehenders can rely on linguistic cues to a great extent (although not completely). However, in oral dialogic discourse, there are conversational cues that go well beyond print, such as intonation, gestures, and the physical environment.

A complete theory of discourse coherence requires a harmonious layering of several levels, including vocabulary, sentence structure, meaning, discourse context, style, and world knowledge. When these levels lack coordination, the coherence is more difficult. To get the message across, the sender will try to coordinate the levels. The receiver assumes that the sender's message is intended to be well formed and will make every attempt to construct a coherent interpretation.

References

Beaugrande, Robert de, and Wolfgang U. Dressler. 1981. Introduction to text linguistics. Harlow: Longman.

Brown, Gillian, and George Yule. 1983. Discourse analysis. Cambridge: Cambridge University Press.

Clark, Herbert H. 1996. Using language. Cambridge: Cambridge University Press.

Costermans, Jean, and Michel Fayol (eds.) 1997. Processing interclausal relationships. Studies in the production and comprehension of text. Mahwah: Lawrence Erlbaum.

Gernsbacher, Morton Ann. 1994. Handbook of psycholinguistics. New York: Academic Press.

Givón, Talmy. 1995. Coherence in the text and coherence in the mind. Coherence in spontaneous text, ed. by Morton Ann Gernsbacher and Talmy Givón. Amsterdam: John Benjamins.

Graesser, Arthur C., Gernsbacher, Morton Ann, and Goldman, Susan (eds.) 2003. Handbook of discourse processes. Mahwah, NJ: Erlbaum.

Graesser, Arthur C., and Gordon H. Bower (eds.) 1990. Inferences and text comprehension. San Diego: Academic Press.

Halliday, Michael A. K., and Ruqaiya Hasan. 1976. Cohesion in English. London: Longman.

Knott, Alistair, and Chris Mellish. 1996. A feature-based account of the relations signalled by sentence and clause connectives. Journal of Language and Speech 39. 143–83.

Mann, William C., and Sandra A. Thompson. 1987. Rhetorical structure theory: toward a functional theory of text organization. Text 8. 243–281.

Van Dijk, Teun A., and Walter Kintsch. 1983. Strategies of discourse comprehension. New York: Academic Press.

MAX M. LOUWERSE AND ARTHUR C. GRAESSER

Color Terms

Color terms have been used in anthropological and cognitive linguistics to investigate whether certain aspects of meaning are universally found in all languages. The classic study of color terminology, Berlin and Kay 1969, is often cited as proof against the Sapir–Whorf hypothesis of 'linguistic relativity' or 'linguistic determinism.' This hypothesis argues that language affects or even determines cognition—that is, that the categories present in the language a person speaks will influence the way that person thinks, the way he or she perceives the world. Languages vary as to which 'basic color terms' occur in the vocabulary. Berlin and Kay set out to test whether the presence or absence of a color category in the language affects the way in which speakers see colors. In the course of their research, they found evidence favoring a 'prototype' theory of meaning and evidence for universal tendencies in the structuring of color vocabulary in the world's languages.

Berlin and Kay gave color charts to speakers of 98 different languages and asked them to sort the 329 colored squares into categories that corresponded to the others): white, black, red, yellow, green, blue, brown, purple, pink, orange, and gray.

Interestingly, these terms occurred in predictable patterns, and Berlin and Kay proposed a scheme of 'implicational universals' to explain the evolution of color terminology. Some languages have only two basic color terms, and these are always *black* and *white*. If a third is added, as in the Arawak language of the Caribbean or in Swahili of East Africa, it is always *red*. After red, languages with four color categories contain either *yellow* or a term that encompasses both blue and green (Kay devised the term 'grue'); languages with five colors include both *yellow* and *grue*, with differentiation between *green* and *blue* occurring in the six-color system. *Brown* is added as the seventh, then *purple, pink, orange,* and *gray,* which do not follow a specific ordering. An example of an implicational universal that follows from this discovery would be that languages with a term for blue will always have words for black, white, red, yellow, and green. The sequence can be depicted as shown below:

black and white	>	red	>	"grue" or yellow	>	yellow or "grue"	>	blue and green	>	brown	>	purple pink orange or gray

basic color terms in their language. They put three limits on what could count as a basic color term:

(1) It had to be a simple word; e.g. *reddish-brown* is not a basic color term.
(2) It had to apply generally, not in a specific domain (like *blonde* or *sienna*).
(3) It could not count as a 'kind of ___', in the way that *emerald* and *olive* are kinds of green.

They found that the most basic color terms any language has is 11, as in English (also Tagalog, Zuni, and

The orderly sequence for the development of color terms leads us to the idea that they might reflect cognitive universals, and this is indeed what Berlin and Kay found in their research on the focal meanings of the color terms. When they asked people to choose the best example of each of the basic color terms from the chart, choices were remarkably similar around the world. For example, people whose language contained no term for orange chose the same focus for yellow as those whose language had terms for both yellow and orange; people whose language contained only three

color terms had the same focus for red as those whose language had all 11 terms. This is the finding that refutes the so-called strong version of the Sapir–Whorf hypothesis (a version espoused by neither Edward Sapir nor Benjamin Lee Whorf), that our perception will be determined by the categories of our language. For color, at least, perception seems to follow universal human neural response patterns in a way that is unaffected by our language habits, although there have been dissenting claims suggesting that Berlin and Kay's experimental methods may not accurately reflect the way speakers apply color terms to objects in the real world.

The existence of universals for the core meanings of color terms does not contradict the fact that culture does play a role both in determining the wider meaning of color terms for speakers and in the inventory of non-basic terms for colors in one's speech repertoire. Color terms may be closely associated metaphorically with objects in our experience, for example, the sun with yellow, or fire with red. Color vocabulary may be highly elaborated in some cultures and for some speakers, according to the importance of color in various arenas of daily life. A painter or an interior decorator, for example, will have an extensive lexicon of color terms. In American society, women have larger color vocabularies than men, due to the sexual division of labor in purchasing clothing and household furnishings and the proliferation of color terms as a form of marketing in American consumer culture.

In applying color terms to objects and to parts of the color chart that are farther away from the focal colors, speakers seem to compare the color in question with the focal color and to judge both similarities and differences. The task of drawing boundaries on the color chart around the limits of each color produced more varied answers than determining the best example of a color did. The focal color seems to act as a prototype for that color. Prototype theory views word meaning as the set of referents that can be designated by that word, which is determined by how those referents compare to a prototypical exemplar of the category. This produces 'fuzzy sets', with some uncertainty about whether referents that are marginal really belong to the category or not. Is *turquoise* a kind of blue or a kind of green? Is *burgundy* a kind of red or a kind of purple? Different people will give different answers.

Researchers continue to debate the ways in which meanings of color terms are encoded in our minds. There are still questions about how language and culture relate to cognition, including whether cognitive images or concepts of color are based solely on neurologically determined perceptions that are universal for humans or whether they are molded in part by linguistic and cultural experience that includes using color words in ways quite different from an experimental setting.

References

Kay, Paul, and Chad K. McDaniel. 1978. The linguistic significance of the meanings of basic color terms. Language 54. 610–46.

Kay, Paul, Brent Berlin, and William Merrifield. 1991. Biocultural implications of systems of color naming. Journal of Linguistic Anthropology 1. 12–25.

Berlin, Brent, and Paul Kay. 1969. Basic color terms: their universality and evolution. Berkeley: University of California Press.

Lucy, John A. 1992. Language diversity and thought: a reformulation of the linguistic relativity hypothesis. Cambridge: Cambridge University Press.

MacLaury, Robert E. 1991. Social and cognitive motivations of change: measuring variability in color semantics. Language 67. 34–62.

Whorf, Benjamin. 1956. Language, thought, and reality: selected writings of Benjamin Lee Whorf, ed. by John B. Carroll. Cambridge, MA: MIT Press.

Wierzbicka, Anna. 1990. The meaning of color terms: semantics, culture, and cognition. Cognitive Linguistics 1. 99–150.

ELLEN JOHNSON

Communication Theory

The contents of the terms 'communication' and 'communication theory' are notoriously vague.

A straightforward psychologist's definition is the following: 'communication is the discriminatory response of an organism to a stimulus'. Thus, Pavlovian conditioning would also be an instance of communication. This is strange to the linguist, because it is generally accepted that not even the learning of languages is a simple conditioning process. How could it be possible that after such a complicated process like language acquisition human beings use their languages for communicative purposes that are as simple as Pavlovian conditioning? The stimulus–response theory does not seem to be on the right track.

Another way to define communication is by information. Often, the term 'communication theory' is used in the strict sense of information theory. Information theory originated in problems around telecommunication techniques and was developed by Claude Shannon and Warren Weaver in 1949. This theory of communication suggests the following schema for communicative acts:

Source ——> Transmitter ——> Channel ——>
Receiver ——> Destination

In effect, the source sends some message to the destination. But let us look at the details between source and destination. First, the message is encoded into a signal; this is done by the transmitter. This point is important in telecommunication; however, in ordinary face-to-face talk, source and transmitter coincide. Leaving the receiver, the signal is sent through the channel. In ordinary face-to-face talk, the channel would be the air channel. But in telecommunication, it would be the electric channel. In the channel, signals are usually changed, or, rather, distorted by the noise. In ordinary face-to-face talk, the surrounding sounds that are not part of the conversation would be the noise of the channel. After the channel, the signal reaches the receiver. There, the decoding of the signal takes place, and we end up with a message again. This message is finally stored in the destination. In ordinary face-to-face talk, receiver and destination coincide.

As precise as this schema is, and as useful as it is for telecommunication, it neglects several factors of natural communication. First, the message lacks interpretation. It seems odd to say that the message is finally just 'stored' in the destination. This is not even true for the simplest acts of communication. If my neighbor at the dinner table utters 'could you please pass me the salt', I would not just store this message in my brain. Instead, I would interpret the message as nonliteral, and furthermore as an instruction to pass the salt to my neighbor. This simple example demonstrates that, often, people communicate because they want others to do certain things. This leads to the next factor of communication that is neglected in the Shannon–Weaver model: the dialogic nature of communication. Natural communication is not just one way as in the Shannon/Weaver schema; normally, something goes back to the sender.

Considerable research has been done to capture the necessary dialogism of communication. The most interesting schools are the Bakhtin Circle, the Prague school, and some versions of functionalism.

Let us focus on the Bakhtin Circle. In the pre-Stalin era, a group of three scholars emerged with some common views on communication and cognition: the literary critic Mikhail Bakhtin, the psychologist Lev Vygotsky, and the philosopher Valentin Voloshinov. Due to political reasons, it took some time before their work reached western research. But since the 1960s and 1970s, their ideas have influenced western psychology and text linguistics. The main idea of these scholars is that the language sign is inherently dialogic, and that all manifestations of the signs are inherently dialogic. Language is formed in the process of social interaction. While the ruling stratum tries to posit a single discourse as exemplary, the subaltern classes are inclined to subvert this monologic closure. In the sphere of literature, poetry and the epic represent the centripetal forces within the cultural arena while the novel is the structurally elaborated expression of popular 'ideologiekritik' (ideology criticism). In Bakhtin's now famous study *Problemy tvorchestva Dostoevskogo* ('Problems of Dostoevskii's work'), it is argued that Dostoevskii's work is imbued with a profoundly democratic spirit. Bakhtin argues that Dostoevskii's creative method is not Hegelian. In a Hegelian scheme, two positions struggle for ascendancy but are transformed into a synthesis at the end. According to Bakhtin, Dostoevskii does not merge the voices into one final, authoritative voice as in the Hegelian absolute. Instead, Dostoevskii presents an unmerged dialogue of voices, each given equal rights. Bakhtin calls this type of dialogue 'polyphonic'. The voice of the narrator resides beside the voices of the other characters. Voices intersect and interact, mutually illuminating their ideological structures, potentialities, biases, and limitations. Polyphony is also observed across different texts: for instance, literary texts show 'intertextuality', which means that they refer to other texts, thus yielding a huge larger text together. 'Voicing' is not only a characteristic of literary texts. In ordinary speech, several factors also lead to a multivoiced message: the cultural context, the social relations between the persons involved in the communication, the context of the situation — all these factors have an effect on the message. Linguistics, however, usually concentrates on only some aspects of communication, leaving cultural and social factors aside.

So far, we have concentrated on the two forms of verbal communication: written and oral communication. But human communication may also be nonverbal. There are many different channels of nonverbal communication: facial expressions, hand gestures, body movements ('kinesics'), and touch ('haptics'). All these modes of nonverbal communication make use of signs, movement, or touch. In animal communication, there are — in addition to nonverbal and (partly) verbal communication — other ways of communication, namely by smell and taste. Human beings do not make much use of smell and taste in the communicative

domain. Humans receive information about the world through the channels of smell and taste, for instance, while eating. But they do not use smell and taste for communication. Body smells seem to have a communicative function only in sexuality.

We have seen that communication has many aspects, and as many as communication theory has. If the focus of attention is more on social factors, the dialogism of communication will be important; if information structure plays a greater role, the theory will be more technical. Little attention has been paid to integrate the findings on nonverbal communication into the general frameworks existing for verbal communication. This might be an interesting field of study for the future.

References

Infante, Dominic A., et al. 1997. Building communication theory. Prospects Heights, IL: Waveland.

Miller, Katherine. 2001. Communication theories: perspectives, processes, and contexts. Mountain View, CA: Mayfield Publishing.

Neuliep, James W. 1996. Human communication theory: applications & case studies. Boston: Allyn & Bacon.

Phillipsen, Gerry. 1992. Speaking culturally: explorations in social communication. Albany, NY: State University of New York Press.

West, Richard, and Lynn H. Turner. 2000. Introducing communication theory: analysis and application. Mountain View, CA: Mayfield Publishing.

MONIKA RATHERT

See also **Text Linguistics**

Comparative Method

The comparative method is the most important of the various methods and techniques used to recover linguistic history. It is also of interest in language classification, in linguistic prehistory, and in research on distant genetic relationships, determining whether seemingly interrelated languages descend from a common ancestor after all. It is generally recognized that 6,000 languages are currently spoken all over the world (one quarter with fewer than 1,000 speakers), and they are grouped into more than 30 widely recognized families. It is estimated that half of those languages will die in the next century. The comparative method is therefore of much help in clarifying their status and evolution through time.

The comparative method is a set of principles and a methodology for comparing forms from related languages to reconstruct an earlier (proto-) language or an earlier state of a language. The aim of reconstruction is to recover as much as possible of the proto-language from a comparison of the descendant languages and to determine what changes have taken place in the various languages that developed from the proto-language.

The comparative method was developed in nineteenth-century Germany for the reconstruction of Proto-Indo-European and was afterward applied to the study of other language families. It is considered the most outstanding achievement of linguistic scholarship in the nineteenth century and was originally stimulated by the discovery by an English orientalist, Sir

William Jones, in 1786, that Sanskrit was related to Latin, Greek, and German. His famous speech to the Asiatic Society of Bengal marks the beginning of comparative philology.

Furthermore, the year 1816, when the German scholar Franz Bopp presented the linguistic public with language material from Sanskrit compared with some other Indo-European languages, remains a historical date in linguistics, because Bopp was the first to realize that the question of the mutual relations of Indo-European languages was worth becoming the subject of specialized inquiry. Indo-European was posited as a hypothetical ancestor language, and comparative studies were almost exclusively focused on the Indo-European language family.

Six years later, another German scholar, Jacob Grimm, included a systematic survey of the relationships between Germanic consonants and their correspondents in Greek, Latin, and Sanskrit in the second edition of his comparative grammar of Germanic languages. He established the existence of a circular 'soundshift' (*Lautverschiebung*) governing these relationships in the prehistory of Germanic. Grimm observed that when other Indo-European languages such as Latin and Greek have a voiced unaspirated stop (*b, d*), Gothic has the corresponding voiceless unaspirated stop (*p, t*), and that, when other Indo-European languages have a voiceless unaspirated stop, Gothic has a voiceless fricative (*f, θ*). In other words, the voiced stops inherited from Proto-Indo-European

became voiceless, and the voiceless stops became fricatives. His discovery is known in linguistics under the name of *Grimm's Law* (or the Germanic consonant shift), although Grimm himself did not use the term *law.* Grimm's Law was probably first worked out in detail by the Danish linguist Rasmus Rask (whose studies, being written in Danish, were less accessible to most European scholars).

However, linguists found at least two classes of exceptions to Grimm's Law with respect to Indo-European voiceless stops: there are cases in which Proto-Indo-European voiceless stops remain voiceless stops in Germanic and cases in which Proto-Indo-European voiceless stops become voiced stops in Germanic. Grimm's Law was thus corrected in 1875 by the Danish linguist Karl Adolf Verner, who examined words in Vedic Sanskrit and Attic Greek that are related to Germanic words that reflect Indo-European voiceless stops as voiced stops. He discovered that in all such cases, the Proto-Indo-European word had intonational stress after the voiceless stop. Verner understood that what seems to be an exception to sound laws may be governed by some other regularity and that sound changes may be unconditioned or may depend on the environment in which the sound occurs. This was a very important step in linguistics, and ancestral common forms could be reconstructed by using the principle of regular sound change and postulating a number of different sound laws that operated independently in the different branches of the Indo-European family.

The comparative method was often criticized because it is based on a misleading genealogical metaphor. In the 1850s, the German linguist August Schleicher transferred the Darwinian view of evolution to language and introduced the model of the evolutionary 'family tree' to comparative linguistics. As at that time it was customary to concentrate on the reconstruction of what was supposed to be the source of all Indo-European languages, Schleicher ventured an actual reconstruction of a short text, the 'fable of the sheep and the horses'.

During the second half of the nineteenth century, a group of four young scholars known collectively as the *Junggrammatiker* ('young grammarians' or 'neogrammarians') proclaimed the 'absolute exceptionlessness of sound laws', i.e. all changes in the sound system of a language as it developed through time were subject to the operation of regular sound laws. The assumption that sound laws were in principle absolutely regular in their operation was quite generally accepted and had become the landmark of the comparative method.

Realizing that sound changes occur in defined geographical regions that do not coincide with the spread of an official standard language, the German scholar Johannes Schmidt criticized the family-tree theory as unrealistic and misleading. According to his theory, known in linguistics as 'the wave theory' (1872), each sound law has its own territory, and different linguistic changes will spread, like waves, from a politically, commercially, or culturally important center along the main lines of communication. These innovations will not necessarily cover exactly the same area.

A real achievement of the method of reconstruction was the Laryngeal Hypothesis presented by the famous Swiss linguist Ferdinand de Saussure in his 300-page *Mémoire sur le Système Primitif des Voyelles dans les Langues Indo-Européennes* (1879; Report on the Primitive System of Vowels in Indo-European Languages): there had to be two hypothetical sounds for Proto-Indo-European, not attested in any of the Indo-European languages known at that time, but sufficient to solve several puzzling irregularities in the development of certain vowels in a number of languages. He was fairly conservative about claiming what they must have been, but he called them laryngeals, indicating that these sounds must have been articulated with the larynx (voice box) as the main articulator, and pointed out the precise locations in words where they must have occurred. Fifteen years after Saussure's death, when the ancient Anatolian language Hittite, the oldest attested Indo-European language, was deciphered, linguists found with amazement that Saussure was right: Hittite had the laryngeal sounds, and they were placed exactly where Saussure had predicted just on the basis of careful reconstruction.

The history of comparative philology (this was the name for linguistics at that time) shows that the comparative method addresses directly only material in the related languages that is inherited from the proto-language and has no means of its own for dealing with borrowings. All things considered, there are four basic assumptions on which the comparative method is based: (1) the proto-language was uniform, with no dialect variation; (2) language splits are sudden; (3) after the split up of the proto-language, there is no subsequent contact among the related languages; and (4) sound change is regular.

Nowadays, the work of reconstruction usually begins with phonology, i.e. with an attempt to reconstruct the sound system; this in turn leads to reconstruction of the vocabulary and grammar of the proto-language. In the nineteenth century, comparative studies were entirely word based: it was words or lexical word stems that were reconstructed and followed through history. Only after the knowledge of phonetics

had begun to increase did linguists realize that they should start reconstruction with sounds, not with lexical items. Consequently, historical linguistics has often enjoyed great success in reconstructing ancestral phonological systems and vocabularies. But for grammatical systems, the situation is slightly more complicated: grammatical patterns are often seriously perturbed by independent and cointeracting processes of change, and grammar usually undergoes more dramatic changes than the sound system and even the vocabulary.

Generally, studies in the comparative method start with the assumption that certain languages are related and hence descended from a common ancestor. The first step in the work of reconstruction is to find sets of putatively related words in the languages or dialects being compared, i.e. placing side by side a number of words with similar meanings from the languages under comparison, for example, Sanskrit *pita,* Latin *pater,* Greek *pater,* and Gothic *fadar*, all of them meaning 'father'. Such sets of 'cognate' forms should be examined for what seem to be systematic correspondences in their sound structure. Each correspondence must yield a plausible-looking sound in the ancestral language, one that could reasonably have developed into the sounds that are found in the several daughter languages; phonological change should be taken into account, because any difference in the form of cognates is a consequence of sound change in one or more of the languages. Establishing regular correspondences between languages permits the assumption that the languages in question are genetically related; these correspondences can also be used to formulate likely hypotheses about what changes gave rise to those correspondences, and to reconstruct antecedent unattested forms in the proto-language from those changes.

A set of 'reconstructed' sounds can then be postulated (marked with an asterisk by the standard convention) to which the sounds in the attested languages can be systematically related by means of sound laws. With these results, each word surviving in the various daughters may be put face to face with its form in the ancestral language. The reconstructed Proto-Indo-European word for 'father' is *pəter*. If neither chance nor borrowing can account for observed similarities, shared inheritance (i.e. 'genetic' relatedness) is the only possibility.

Languages replace lexical items over time; the more the time, the fewer the items retained from the ancestral language, and the fewer the potential cognates. This is why linguists agree that it is convenient to start with cognates belonging to the basic vocabulary, because they resist borrowing more than other components of the vocabulary. The final step is to determine what system of sounds the ancestral language apparently had and what the rules were for combining these sounds.

The success of any reconstruction depends on the material at hand and on the ability of the comparative linguist to decipher what happened in the history of the languages under comparison. Generally, the longer in the past the proto-language split up, the more linguistic changes will have occurred and the more difficult the task of reconstruction with full success is going to be.

The choice of languages is also very important. If we simply pick some arbitrary languages and compare them, we cannot expect to see systematic correspondences appear, even if the languages selected are indeed genetically related. The comparative method has to be applied carefully and thoughtfully, and every piece of information must be taken into consideration as possibly relevant. Linguists have to be aware of a number of potential difficulties that might lead the reconstruction into error. Perhaps the most obvious point is that the comparative method cannot recover any feature of the ancestral language that has disappeared without a trace in all the attested daughters. Other problems are loanwords, pure coincidences, nursery words, imitative words, and phonaesthetic words (sound symbolism), because these tend to be somewhat similar even in unrelated languages.

References

Beekes, Robert S.P. 1995. Comparative Indo-European linguistics: an introduction. Amsterdam, Philadelphia: Benjamins.

Brogyanyi, Bela, and Reiner Lipp (eds.) 1993. Comparative-historical linguistics: Indo-European and Finno-Ugric. Amsterdam, Philadelphia: Benjamins.

Collinge, N.E. 1985. The laws of Indo-European. Amsterdam, Philadelphia: Benjamins.

Meillet, Antoine. 1953. Introduction à l'étude comparative des langues indo-européennes, 8th edition. Paris: Hachette.

Meillet, Antoine. 1967. The comparative method in historical linguistics, translated from the French by G.B. Ford, Jr. and H. Champion Paris. Oslo: Instituttet for Sammenlignende Kulturforskning.

Saussure, Ferdinand de. 1968. Mémoire sur le système primitif des voyelles dans les langues indo-européennes (1879; Report on the Primitive System of Vowels in Indo-European Languages) (Reprografischer Nachdruck der Ausg. Leipzig 1879.) Hildesheim: Olma.

Sihler, Andrew. 1995. New comparative grammar of Greek and Latin. New York: Oxford University Press.

LAURA DANILIUC

See also **Bopp, Franz; Grimm, Jacob; Historical Linguistics; Saussure, Ferdinand de**

Compositional Semantics

Compositional semantics includes all approaches to the meaning of natural language that subsume, in one formulation or another, the principle of 'compositionality of meaning'. Very generally put, this principle reads: 'The meaning of a complex expression is a function of the meanings of its parts.' This is often called Frege's principle, although it cannot be found in explicit form in his writings. Almost any theory of meaning for natural language is based on this principle. There is a psychological motivation for this, because the principle is able to explain how a speaker can understand sentences never heard before.

A more explicit formulation of the principle is: 'The meaning of an expression is a function of the meanings of its parts and of the way they are syntactically combined.' This formulation makes it explicit that the meanings of the parts are not sufficient for determining the meaning of the whole, and that the way the parts fit together is responsible for differences in meaning (*John hit Mary* vs. *Mary hit John*). In this form, the principle is rather uncontroversial, although when one looks at the way the relation *is a function of* is spelled out in different semantic theories, there is considerable diversity.

This principle is the fundamental one in Montague Grammar, an application of the essential methods of formal logic to the study of natural language semantics: it helps in designing a finite system (with a limited number of words and ways of combining them) with infinite output (meanings for all possible sentences). One of the basic tenets of Montague Grammar is the 'rule-to-rule' correspondence between syntax and semantics: the syntax contains several rules that provide several ways to form a complex expression from parts, each with a specific semantic effect (i.e. for each syntactic rule, there is a corresponding semantic rule). Put informally, the meaning of a complex expression results from the application of a semantic rule to the meanings of the parts of the expression.

Some natural language expressions have been claimed to be difficulties for a compositional approach to semantics. One such difficulty is represented by compound nouns such as *bookworm* (which is not normally a worm but a person) or *ashtray* (which is not a tray). However, most of these challenges have been successfully met by insightful compositional analyses.

Due to its psychological plausibility, the compositionality principle characterizes not only formal approaches to semantics but also one major area of cognitive linguistics. In Adele Goldberg's Construction Grammar, for instance, syntactic constructions are considered to be substantive principles of semantic composition: since the constructions per se are taken to be meaningful, the meaning of linguistic expressions is the result of integrating the meanings of the words into the meanings of constructions.

References

Goldberg, Adele E. 1995. Constructions. Chicago, IL: University of Chicago Press.

Janssen, Theo M.V. 1997. Compositionality. Handbook of logic and language, ed. by Johan van Benthem and Alice ter Meulen. Oxford: Elsevier and Cambridge; MA: MIT Press.

Katz, Jerrold J., and Jerry A. Fodor. 1963. The structure of a semantic theory. Language 39.

Partee, Barbara H. (ed.) 1976. Montague Grammar. New York: Academic Press.

Partee, Barbara H. 1984. Compositionality. Varieties of formal semantics, ed. by Fred Landman and Frank Veltman. Dordrecht: Foris.

Thomason, Richmond H. 1974. Formal philosophy. Selected papers of Richard Montague. New Haven, CT: Yale University Press.

ANDREA SANSÒ

See also **Meaning; Montague, Richard; Semantics**

Compounding

Compounding is the process of creating new words whose elements are smaller words. For example, the relatively new English word *chatline* is created by compounding (a process sometimes termed 'composi-tion') from the two simpler words *chat* and *line*. A word formed in this way is called a 'compound'.

The 'words' that are the constituent parts of compounds may be subject to specific formal requirements,

which depend on the language concerned. In the Latin compound *liquefacio,* 'to dissolve', *facio* means 'I make', but *lique,* although clearly related to *liquidus,* 'liquid', cannot occur by itself, but only in compounds and other derivative word-forms. In the Danish compound *sovesofa,* 'sleep sofa, sofa bed', the verb in the first element of the compound is in the infinitive. In other languages, although it is rare, finite verb-forms can be used, as in Arabana-Wangkangurru *yanhi·rnda· tharka·kura* 'speak-pres-stand-past.continuous = they were standing talking'. In languages like Finnish and Sanskrit, both basic and inflected forms of nouns can be found in the first element of compounds. Compounds usually consist of two parts: a modifying word and a modified one. The modified word is called the 'head' of the compound because it contributes the dominant meaning and determines whether the compound is, e.g. a noun or a verb. Whatever the form of the modifying element of the compound, it is usually fixed in the compound and does not change depending on the environment in which the compound occurs. The head element, however, may be inflected, e.g. for tense (in the case of verbs) or number (in the case of nouns).

Typically, compounds denote hyponyms, or special cases, of their head. A *chatline*, for example, is a type of line, not a type of chat. In English, the head is almost always the right-hand element in the compound. In a language like French, however, the head is frequently on the left: a *leçon-cuisine,* literally 'lesson cookery' (cooking lesson), is a type of *leçon* and not a type of *cuisine*. *Rolls-Royce* is quite correctly translated into French as *Royce-Rolls.* A remarkable number of languages show both patterns of headedness in compounds. In either case, compounds are ideal constructions for creating names for subtypes.

Because compounds are typically hyponyms of their head element, they also take their part of speech from their head element: for example, *chatline* has a noun in its head and is hence a noun, *user-friendly* has an adjective in its head and is thus an adjective, and *stage-dive* has a verb in its head and is consequently a verb.

Compounds in which this is true are called 'endocentric compounds' because their center of meaning lies within the compound itself (Greek *endo,* 'inside'). There are also 'exocentric compounds' (Greek *exo,* 'outside'), which are not hyponyms of their head element (an *egghead* is a kind of person, not a kind of head) or, more radically, whose part of speech is not deducible from that of their elements (for example, *put-down* consists of a verb, *put,* and a preposition, *down,* but it acts as a noun).

Although exocentric compounds are not hyponyms of their own heads, they can still be seen to be headed, in that one element modifies the other. A *redneck* may not be a type of neck, but it is still clear that *red* modifies *neck* and that *redneck* thus has its own right-hand head;

a *bas-bleu* (French) 'bluestocking' (literally, 'stocking blue') may not be a stocking, but not only does *bleu* obviously modify *bas*, the construction as a whole takes its gender from *bas*.

The same is not so obviously true of another type of compound, sometimes called 'coordinative compounds'. Here, two elements are put together, and the compound either denotes the unity of the two elements or is used as the superordinate term for things that include the two named elements. In the first case, we get things such as Marathi *āī-bāp,* 'parents' (literally 'mother–father'); in the latter case, we get things like Kannada *bassu karu,* 'vehicles' (literally 'bus–car'). The borderline between coordinative and endocentric compounds is often blurry, and classifications may vary for different scholars: several distinguishable classes of compounds similar to coordinative compounds are found, but they are not always kept clearly apart.

All the examples that have been considered so far are examples of what are usually called 'root' or 'primary' compounds. One of the typical characteristics of primary compounds is the wide range of possible meaning relations between the elements. For example, a *flour mill* grinds flour, but a *water-mill* is powered by water, whereas a *sawmill* uses a saw as an instrument in its work. There is nothing overt in these words to show that they have different interpretations, and the same form may have different interpretations on different occasions: the *Washington flight* may be the flight to Washington or the flight arriving from Washington.

A contrasting set of compounds, usually called 'synthetic' or 'verbal' compounds (some scholars distinguish between the two labels), has a very restricted interpretation. These are words such as *bus driver*, in which the head element is a derivative of a verb and in which the modifying element must be interpreted as an argument of that verb. The argument concerned is usually the direct object of the verb (as in *bus driver,* 'someone who drives buses') but may on occasions be other arguments, such as instrument in *hand-sewn* and subject in *consumer spending.*

Precisely where compounding stops and sentence formation begins is a matter of some dispute. In particular, it is not clear whether structures like Onondaga *waʔ·ha·hwist·ahtu·ʔt·aʔ,* 'he lost money' (in which the direct object *money* is directly attached to the verb *lost*) should be regarded as a compound or some other construction.

References

Adams, Valerie. 1973. An introduction to modern English word-formation. London: Longman.

Bauer, Laurie. 2001. Compounding. Language typology and language universals, ed. by Martin Haspelmath, Ekkehard

König, Wulf Oesterreicher, and Wolfgang Raible, 695–707. Berlin and New York: de Gruyter.

Botha, Rudolf P. 1984. Morphological mechanisms. Oxford: Pergamon.

Coates, Richard. 1999. Word structure. London and New York: Routledge.

Ryder, Mary Ellen. 1994. Ordered chaos. The interpretation of English noun–noun compounds. Berkeley, CA: University of California Press.

Williams, Edwin. 1981. On the notions 'lexically related' and 'head of a word'. Linguistic Inquiry 12. 245–74.

LAURIE BAUER

Computational Linguistics

Computational linguistics is a discipline between computer science and linguistics, for which the main concern is the computational aspects of human language. These are aspects that can be put into the form of a sequence of instructions that a computer can understand. For example, if we have a sentence like *John runs*, we can set up some simple grammar rules. These rules state that a sentence is made up of a noun and a verb, a noun is always *John,* and a verb is always *runs*. Of course, this grammar only produces our sentence, but it is a grammar that can be used very easily in a computer program. We only have to put three instructions (Sentence = Noun + Verb, Noun = *John*, and Verb = *runs*), in a format that the machine understands, to produce the sentence. We say then that this grammar is fit for computation. We do not include in our definition those tools that computer science gives us for the study of language. Sometimes these are also presented under the generic label, but they are normally not considered to be part of this field.

Two main practical goals in computational linguistics have been *machine translation* and *natural language interaction* with computers and machines. When we refer to 'natural language', we are doing so in contrast to 'artificial language'. Language can be called 'natural' when it can be used for communicating with people in the same way as a human being would do.

Machine translation, that is, translation carried out by a computer, was the first goal of computational linguistics. When scientists started to work in this field, it looked like an easy task. It was thought to be just a question of making words and expressions in the source language correspond to words and expressions in the target language. Very soon it became clear that it was not so simple, since language has a structure that is more complex than just a lexicon database. There is more to language than just words.

However imperfect the process still is, though, automated translation now gets much better results than in the past. And it has achieved them through a systematic approach to how to perform the task. From the very early years of machine translation studies, we have two main approaches to the problem of automating translation. The first method involves carrying out a direct translation from one language into the other. The second method is the 'transfer' grammar approach, in which everything is translated to a kind of intermediate language. From here, it is easier to make a translation to many other languages. This intermediate grammar may have the form of a real human language (Esperanto or English are examples), or a certain symbolic language. A simple example is the following: we have the French sentence *Marie aime Jean* and then we can make a translation to a simplified English sentence of the form *Marie(S)-love-Jean(O)*, with *S* marking the subject and *O* the object. From this intermediate sentence, we translate again, now into Spanish: *Marie ama a Jean*. The technique is very useful when translating to and from many different languages, since we have an intermediate expression from which we can easily make translations that are faithful to the same extent to the basic structure. Both techniques have their advantages and drawbacks.

In addition, even if there should not be any problem at all with the translation strategy adopted, there are other facts that machine translation has not properly accounted for to date, the most important being the resolution of ambiguity in language. The phrase *The cover of the book which I like* has syntactic ambiguity, since we cannot be sure whether what I like is the cover or the book. Similarly, *The seal is on the rock* has semantic ambiguity, since we do not know whether what is on the rock is an animal (*a seal*) or some kind of stamp (*a seal*). We may also have pragmatic ambiguity, of which a very simple case is the difficulty to translate the English pronoun *you* into languages like Spanish, French, German, or Japanese, where there are several different ways of referring to a second person according to degrees of formality or to the relation between the speaker (or writer) and the

hearer (or reader). Obviously, the computer cannot infer this type of information, which depends on the communicative situation.

Of all these types of ambiguity, the most important for machine translation is the last one, since it cannot be resolved by looking at the immediate linguistic context.

Another main area that computational linguistics has always been interested in is the interaction with machines using natural language. This can be done through the use of the written word or through the use of oral speech. In the latter case, the system needs to incorporate a module for speech recognition, in addition to the other modules used for linguistic processing that we have mentioned. The usefulness of mechanisms that react to ordinary speech in activities where free hands are needed has led to the development of such systems for cars, planes, or control of processes in factories, which are typical examples. In all these cases, in addition to the procedures of language analysis and generation, there are special calculations that take into account the noise that may accompany ordinary language.

In particular, there has been a special concern in trying to produce language in a manner as human-like as possible in *expert systems*. An expert system is a system capable of reasoning in a way similar to a specialist in a certain field, and it is therefore necessary, or at least very convenient, that the queries can be carried out in a language as close as possible to natural human language. This is so because expert systems are intended for use in many kinds of settings, and may have to be operated by persons with different degrees of expertise. It is also important to try to reproduce the typical interactive client–expert exchange.

In natural language interaction there are two main strategies. The oldest and the easiest is to check for certain words and expressions, ignoring the rest of the message. This is the direct, crude approach. According to the expressions the system recognizes, it gives certain answers. This was the approach used in Eliza, a program created by Joseph Weizenbaum in 1964–1966. It was designed to take on the role of a Rogerian psychotherapist, and it produced responses according to the trigger expressions that appeared. Thus, to a sentence with the word *friends*, it could react saying something like *Why are you thinking of your friends now?* so that to the sentence *I have many friends*, or *My wife has many friends*, we would get the question, *Why are you thinking of your friends now?* The rigidity of this approach is obvious, although it is the most economical by far.

On the other hand, we have systems where there is a special module for language recognition, in which sentences are *parsed* (analyzed) and semantic inter-pretations are assigned to them. This is the true computational linguistics approach, which has also found problems similar to the ones mentioned for machine translation. It is this approach that shall be explained below.

In computational linguistics, the structure of sentences needs to be represented in some way. In order to do this, the computer has some internal arrangement of data that corresponds to the semantic structures of the knowledge that we have in our mind, and that helps clarify the way in which the linguistic message organizes the information it conveys. This arrangement is in the form of a *semantic representation*. The computer also knows about the rules of the language for putting words and sounds (or letters) together. These are *syntactic and phonetic (or graphemic) rules*.

The semantic representations used in computational linguistics try to make the most of the resources that the machine offers, but they also try to be faithful to the structure of concepts. Let us take the example of the concept 'bird' to illustrate how these representations can be made. When we think of a bird, we immediately visualize an animal with wings and feathers. In our mind, we have a link between the concept 'bird' and the concepts 'wing' or 'feather'. If we write the words 'bird', 'wing', and 'feather', and draw two lines from the label 'bird', one to the label 'wing' and another to the label 'feather', we will have a representation of the above image by using a labeled diagram. If we now add more words to the diagram, such as, for instance, 'beak', 'egg', 'nest', or 'to fly', the diagram becomes more complex, but now the links start to be of different kinds. A bird normally 'has' a beak, 'lays' eggs, 'builds' nests, or just 'flies'. These different types of relation also need to be labeled. Finally, we now know that not all these features are necessary and sufficient, but rather prototypical. It is difficult to imagine a bird without a beak or wings, but certainly not all of them fly or build nests. This is more difficult to put in our diagram, but we can try to use some numbers that denote to which extent these features can be applicable.

When our diagram becomes larger, it should look like a *net*. In fact, it is a way of depicting a *semantic net*, a kind of representation that is being used in computational linguistics since the late 1960s. Semantic nets are incorporated in the data of computational linguistic applications, so that the semantic structure of sentences is interpreted or generated by building small nets that match the connections that appear in the larger ones. As a formalism, a semantic net consists of a collection of points (the labels 'bird', 'wing', 'feather', etc. in our example) called *nodes*, each representing a concept. A node may be

connected to any other node in the net by means of *relations* that in turn carry labels ('have', 'build', 'fly', etc.). There are some conventions: two very typical relations normally referred to in computational linguistics books are the labels IS-A (x 'is a' y) and H-A-P (x 'has as part' y). But any other that corresponds to common English verbs or adjectives (LIKE, COLOUR, etc.) can also be used.

Let us look at another example. A semantic net for the sentence *Mary likes Peter* could be the following:

GIRL		BOY
IS-A		IS-A
MARY	LIKES	PETER
H-A-P		
EYES		
COLOUR		
BLUE		

This net provides a semantic representation for the message conveyed. However, most of the information is not in the sentence itself, which only carries labels. We retrieve all the additional information from structures of knowledge that are either in our mind or in the computer's data. There is a matching process in order to see which parts of the sentence's semantic structure and the program's data coincide. It is like having two pictures in two transparencies, A and B, of a face and a nose. The nose has more detail in B, but it coincides partially with what is drawn in A; hence, we know that it corresponds to that face. In addition, when we put the two transparencies together, the face receives a better drawing, with the addition of a more detailed nose.

This matching process is currently referred to in computational linguistics as *unification*, and it is considered a very powerful resource for natural language processing. When a computer programmer uses a programming language such as PROLOG, for instance, there is an extensive use of unification.

We can illustrate this by having a close look at how PROLOG works. As has just been stated, PROLOG is a programming language, and all programming languages have a vocabulary and a syntax, exactly in the same way as natural languages do. In the vocabulary, we have strings of symbols that stand for *variables* (with capital letters, e.g. *X, YW, AA*) or for *constants* (without capital letters, e.g. *a, bb, person, mother*). The corresponding items in a natural language for variables would be pronouns (which stand for other terms). For constants, we have proper nouns (which, in ideal cases, refer to unique entities).

Statements in PROLOG normally have a structure in which there is one predicate and several arguments. If A is the predicate and x, y the arguments, the convention is to write A (x, y). The symbol :- is used to assign several statements as conditions for the truth of another one, and the resulting structure is called a *rule*. An example will help to understand this. Below there is PROLOG rule:

```
aunt (X, Y) :-
    female (X),
    brothers (X, Z),
    parent (Z, Y).
```

Let us translate these statements into understandable English (on the right):

aunt (X, Y) :-	'X (a variable) is Y's (another variable) aunt if...'
female (X),	'X is a female'
brothers (X, Z),	'X and Z (a third variable) are brothers/sisters'
parent (Z, Y).	'Z is a parent of Y'

In plain English, this would be something like 'Some entity X is the aunt of some other entity Y if the entity X is a female, the entities X and Z are brothers or sisters, and the entity Z is one of Y's parents.'

The computer programmer writes this rule as part of a larger program, and we can also add some information to the program's database by writing the following statements:

female (mary)	'Mary is a female'
brothers (mary, john)	'Mary and John are sister and brother'
parent (john, jim)	'John is one of Jim's parents'

(Notice that the terms 'female', 'mary', 'john', etc. have no capital letters, because they are constants.)

The computer operator may wish at some stage to know who is Mary's nephew, or to whom she is an aunt. We can then make use of the 'aunt' rule. The operator writes:

?- aunt (mary, Y)   'Mary is Y's aunt'

As this is a plain statement (not a rule), the PROLOG system looks for a suitable term that can change the variable Y into a constant. If you look at the information provided in the statements that have been incorporated into the program so far, this term is *jim*. Accordingly, it answers:

Y = jim

We then say that *jim instantiates* the variable Y.

The PROLOG system has carried out a *unification* process in order to come to the right solution. It has

unified the rule and the statements in the following way:

```
aunt (mary [=X], Y) :-
    female (mary [=X])
    brothers (mary [=X], john [=Z])
    parent (john [=Z], jim [=Y])
```

So that Y = jim.

Unification is a very important procedure in computational linguistics. It simulates on a computer the processes that take place when we compare information from our mental store with the information in the linguistic message.

Another well-known programming language for artificial intelligence is LISP, of which many different versions have been created afterwards (schema, xlisp, etc.). LISP stands for 'List Processing', and it is a complementary approach to PROLOG. Whereas in PROLOG we had a deductive mechanism that unified knowledge structures, in LISP we have a powerful engine for processing lists. This is useful in computational linguistics, where syntactic parsing is an important step in the analysis of sentences. An additional important feature, which is very convenient for many artificial-intelligence approaches is that LISP programs are able to self-modify, making it possible to write code that 'learns', or that adapts itself to changing circumstances.

PATR (Shieber 1986) is a declarative language that has eventually become a standard in computational linguistics. This is so because it is very easily adaptable to many different theories. In fact, many current developments use feature structures and unification in a PATR style. The most prominent characteristics of the other two languages that we have seen here, unification (PROLOG) and string concatenation (LISP), are used in this language.

With respect to semantic nets, computer scientists soon found that representations like the ones shown above fail to give detailed information about the situations that are represented. Location, time, setting, etc., were notions that did not appear in the structures. This led to more complex representations, in which concept nodes were used for events as well as objects, and cases appeared as labels in semantic nets. Parallel to the relation IS-A for objects, there is also now ACT-OF. In the sentence *John throws the ball*, John is linked through an AGENT relation to a node. 'The ball' is also linked to that same node by the relation OBJECT. And the node is linked to 'throw' by means of ACT-OF, which therefore links particular instances of events to the general node for a certain action in the knowledge database.

An example of how the semantic net paradigm can be realized is R.C.'s (1973) theory of conceptual dependency. This author's graphs follow the formal structure of a dependency grammar, and have four different types of nodes. These are symbolized by PP, ACT, PA, and AA, corresponding very roughly to noun, verb, adjective, and adverb, respectively. Schank makes use of a small set of primitive actions (PROPEL, MOVE, INGEST, EXPEL, GRASP, PTRANS, MTRANS, ATRANS, SPEAK, ATTEND, and MBUILD) and the default action DO. With the help of a small number of states, this author claims that the meaning of all verbs can be represented. It is interesting to see how the semantic primitive approach is used here. In addition, we also have rules that must be applied for well-formed conceptualizations.

Sets of semantic primitives have been used in many CL systems. However, it is important to clarify here that there is no claim about their universality or a very strong psychological justification. The use of a limited set has more to do with a question of convenience.

References

Schank, R.C. 1973. Conceptual dependency: a theory of natural language understanding. Cognitive Psychology 3. 552–631.

Shieber, Stuart M. 1986. An introduction to unification-based approaches to grammar. CSLI Lecture Notes, Vol. 4. Stanford.

Wilks, Yorick. 1973. Preference semantics. Memoranda from the Artificial Intelligence Laboratory 206. Stanford, CA: Stanford University.

CARLOS INCHAURRALDE

See also **Ambiguity; Artificial Intelligence; Machine Translation**

Computer-Assisted Language Learning (CALL)

Computer-assisted language learning (CALL) has been an important aspect of language teaching and learning since the 1960s, and its developments have been influenced and shaped by those occurring more generally in the fields of education and applied linguistics. The first applications of computers in language teaching were

based on large-scale, specialized mainframe computers, and hundreds of person-hours of programming were needed to develop the instructional software. These computer-assisted instruction (CAI) systems were meant to be 'stand-alone', which meant that all aspects of teaching and learning were contained within the hardware and software. It was very difficult (or often impossible) for teachers to tailor the software to their classroom teaching, and some of the earliest complaints about CAI were related to this problem. The teaching approach used in CAI systems was tutorial (drill and practice) and was based on behaviorist learning theories that were popular at the time. Among the early CAI projects, perhaps the two most influential for language teachers, were PLATO (Programmed Logic for Automatic Teaching Operations) and TICCIT (Time-shared, Interactive Computer-Controlled Information Television), developed at the University of Illinois and Brigham Young University, respectively.

In the 1970s and 1980s, the change to more communicative approaches to language teaching and learning coincided with the development of microcomputers and the corresponding proliferation of software. Many programs were still tutorial (drill and practice), but more general learning tools, including, for example, word-processing, database, spreadsheet, and, by the 1980s, communications software, were adopted from the business environment. In addition, authoring software (high-level programming languages) helped teachers write computer programs to complement their other classroom activities. Consistent with new communicative approaches to language teaching, teachers began to group students so that they could work together in teams to write simple software programs. Not only did this new instructional approach help students to become more independent learners, but also team-programming provided them with opportunities to practice language in authentic communicative contexts.

Since the 1990s, with the advent of inexpensive memory and other hardware components, for example, resident CD-ROM or DVD (digital video disk) players, teachers have had access to a range of commercially developed, multimedia language software. Such software is usually created as a hypertext environment in which text, sound, video, and graphics are creatively linked to one another. Students can point and click on different words, ideas, or branches of a program and move through instructional material at their own pace, according to their personal interests. However, multimedia software can be expensive and cannot always be tailored to individual classroom learning needs. Some researchers have also questioned whether reading and writing in hypertext environments, disjointed as it can be, helps or hinders language learning.

Also since the early 1990s, access to global communication networks and the World Wide Web, and the availability of point-and-click Web-browsing software, has become increasingly easy to use and inexpensive. The Internet has made global communication, through electronic mail (e-mail), commonplace and has led to a variety of tandem language-learning projects. For example, English-speaking students studying French in the United States could be paired electronically with French-speaking students studying English in Canada. Both groups write to each other in the target language and then send, read, or reply to messages at their convenience; they do not need to be on-line at the same time (asynchronous communication). Learning is aided as students work at their own pace and help each other to notice problems in grammatical forms or learn how to express ideas in more natural, idiomatic language. In addition, students can provide insights to their e-mail partners about respective cultural practices and norms. Such peer-based sharing of ideas to help learners develop a deeper understanding of language is called 'scaffolding' and has been influenced by learning theories from education and psychology.

In contrast, synchronous communication, in which all people are on-line and writing (typing) at the same time, is also used in CALL classrooms. Synchronous communication, for example, in on-line chatrooms, requires students to log into centralized 'rooms' on the Web, where they can adopt pseudonyms or even assume different on-line identities. The pace of written exchanges is fast, and the style of language is often informal. Many language teachers believe that unless there is a well-developed curriculum structure and educational purpose to synchronous communication, students may not learn appropriate linguistic forms. In fact, some researchers believe that on-line chatroom communication involves a new, hybrid form of language that incorporates features of both speaking and writing.

Although the Internet supports the exchange of written text through e-mail or in chatrooms, it also provides access to a variety of other language-teaching resources. These include, for example, access to Web-based software (much of which can be downloaded) and specialized Web sites containing organized links to language-learning resources around the world. For example, students can read daily newspapers, listen to live radio broadcasts, or watch and listen to television news from a variety of countries. As a result, access to 'authentic' language resources is becoming easier for teachers, but how to evaluate, organize, and integrate on-line materials so that students can learn language in a structured manner have become important issues in teacher training. Certainly, there is a growing acceptance

that the role of the teacher and student in CALL class-rooms is changing, with teachers assuming more facilitative, guiding roles and students becoming more autonomous learners.

The latest technological changes that are affecting CALL include wireless communication, which relies on radio signals, rather than physical wiring, to connect to the Internet, and a variety of new software tools. For example, it is now possible for students to use wireless laptop computers, video cameras, and word-processing and movie-editing software to write, film, and review their own language materials. Such a project-based approach to teaching and learning could effectively integrate current theories of autonomous learning, peer scaffolding of learning, and authentic communication. Regardless of the technology used, however, the overriding issue for teachers will remain how to coordinate learning experiences so that students have meaningful access to language—however sophisticated their computers and whichever software they use.

References

Bush, Michael D., and Robert M. Terry (eds.) 1997. Technology-enhanced language learning. Lincolnwood, IL: National Textbook Company.

Egbert, Joy, and Elizabeth Hanson-Smith (eds.) 1999. CALL environments: research, practice, and critical issues. Alexandria, VA: TESOL Publications.

Levy, Michael. 1997. Computer-assisted language learning: context and conceptualization. Oxford: Clarendon Press and New York: Oxford University Press Inc.

Murray, Denise E. 1995. Knowledge machines: language and information in a technological society. New York: Longman.

Pennington, Martha C. 1996. The power of CALL. Houston: Athelstan Publications.

Taylor, Todd W., and Irene Ward (eds.) 1998. Literacy theory in the age of the internet. New York: Columbia University Press.

Warschauer, Mark. 1999. Electronic literacies: language, culture, and power in online education. Mahwah, NJ: Lawrence Erlbaum Associates.

Warschauer, Mark, and Richard Kern (eds.) 2000. Network-based language teaching: concepts and practice. Cambridge: Cambridge University Press.

E. MARCIA JOHNSON

Comrie, Bernard

Bernard Comrie is one of the leading figures in language typology. As a fieldwork typologist, Comrie has carried out studies on Fula, Tunisian Arabic, Malayalam, Moroccan Arabic, Khmer, Tsez, Sudanese Arabic, Maltese, and Basque, on which he has taught field methods classes. He has also worked in depth in the field on two additional languages: Haruai and Bezhta. Comrie has also taught and researched Historical Linguistics and Diachronic Syntax.

Although typological linguistics is the overall approach that is closest to his own, his work has very close ties with cognitive linguistics. Even if as a formal framework, Relational Grammar is now virtually dead, in terms of formal approaches to grammar, Comrie believes that linguists have learned a lot about the grammars of different languages and about the general properties of language from that framework.

Comrie has been influential in the recent upsurge of interest in linguistic typology, both through his general books on the topic and through individual articles. Within typology, the areas in which he has had most impact are relative clauses, causative constructions, and grammatical relations. In addition, his books on tense and aspect have had an important role in generating interest in these topics, even among linguists who end up disagreeing with his approach.

Some common linguistic terms are actually used for the first time in his papers. The term 'accessibility hierarchy' came with Keenan and Comrie (1977). His book *Aspect* played a major role in getting the distinction widely accepted between 'perfect' and 'perfective'. He is perhaps the first linguist to introduce, among other terms, the 'S, A, P' terminology (as opposed to Dixon's similar 'S, A, O').

One of the main features in his thinking has been the integration of links between language and other phenomena, in particular cognitive and social. He is in search of correlating linguistic and other (genetic, archeological) evidence for human prehistory. He insists that these different disciplines provide us with different windows into human prehistory that must then be reconciled with one another, rather than presupposing, for instance, a close connection between language families and biological populations.

Comrie's ideas have not remained static. In some of his later articles (e.g. 'Rethinking the typology of relative clauses', 1998), he has argued for substantial revisions to the ways in which typologists and other

linguists look at relative clauses, in particular, paying more attention to semantic and pragmatic factors. With regard to aspect, he gives more import to the role of lexical properties of verbs than he did in the mid-1970s.

Comrie's first publications were articles derived from his doctoral dissertation, and thus dealing with generative grammar. However, his first articles in the field of universals and typology include 'Noun phrase accessibility and universal grammar', written jointly with Edward L. Keenan (1977). He has a remarkable number of publications, mostly on the typology of ergativity, aspect, causative constructions, and nominalizations. His major publications include *Aspect* (1976), *Lingua descriptive studies: questionnaire* (with Norval Smith, 1977), *The Russian language since the Revolution* (with Gerald Stone, 1978), *Language universals and linguistic typology: syntax and morphology* (1981), *The languages of the Soviet Union* (with B.G. Hewitt and J.R. Payne 1981), *Tense* (1985), and *The Russian language in the twentieth century* (with Gerald Stone and Maria Polinsky 1996). To this list, we should add his several important edited volumes (like *Classification of grammatical categories* (1978), *Studies in the languages of the USSR* (1981), *The world's major languages* (1987), *The Slavonic languages* (with Greville G. Corbett 1993), and some others, along with dozens of articles on language typology.

Biography

Bernard Comrie was born on May 23, 1947 in Sunderland, England. By 14 years of age, he knew French and had already started learning Latin, German, Ancient Greek, and Russian. Coming across R.H. Robins' textbook on linguistics left a deep impression on him, showing him not only that there is a field called linguistics, but that it is also a systematic, scientific endeavor.

Comrie studied Modern and Medieval Languages (mainly French, German, Russian, Linguistics) at the University of Cambridge, receiving a B.A. (1968) and a Ph.D. in Linguistics (1972). His dissertation 'Aspects of sentence complementation in Russian' was written under the supervision of Pieter A.M. Seuren.

Comrie taught linguistics as a University Lecturer in Linguistics at the University of Cambridge (1974–1978). In 1978, he joined the Department of Linguistics at the University of Southern California, Los Angeles, and worked there as a Professor of Linguistics until 1998. Since 1997, he is director of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, in addition to being a part-time Distinguished Professor of Linguistics at the University of California, Santa Barbara (since 2002). In addition, he holds honorary positions as a Research Professor of Linguistics, University of Southern California (since 1998) and as a Professor of Linguistics, University of Leipzig (since 1999).

References

Comrie, Bernard. 1976. Aspect. Cambridge: Cambridge University Press.

Comrie, Bernard. 1979. Noun phrase accessibility and universal grammar. Linguistic Inquiry 8. 63–99.

Comrie, Bernard. 1989. Language universals and linguistic typology: syntax and morphology, 2nd edition. Oxford: Blackwell and Chicago: University of Chicago Press.

Comrie, Bernard. 1998. Rethinking the typology of relative clauses. Language Design 1. 59–86.

Dahl, Östen. 1987. Comrie's *tense*. Folia Linguistica 21. 489–502.

Declerck, Rennat. 1986. From Reichenbach (1947) to Comrie (1985) and beyond: towards a theory of tense. Lingua 70. 305–64.

Kulikov, Leonid, and Heinz Vater (eds.) 1998. Typology of verbal categories: papers presented to Vladimir Nedjalkov on the occasion of his 70th birthday (Linguistische Arbeiten 382). Tübingen: Max Niemeyer.

Rakhilina, Ekaterina V., and Yakov G. Testelets (eds.) 1999. Typology and linguistic theory: from description to explanation, for the 60th birthday of Aleksandr E. Kibrik, 319–39. Moscow: Jazyki russkoj kul´tury.

BEHROOZ MAHMOODI-BAKHTIARI

Configurationality

English is a configurational language because the grammatical functions subject and object appear compulsorily in particular syntactic positions: nearly every English clause must have an overt preverbal subject and every English transitive clause must have an overt postverbal object. Of the three main constituents of the clause—subject, verb, and object—the syntactic relationship between verb and object is *bounded*, as opposed to the syntactic relationship between subject and verb, which is *unbounded*. The strong syntactic

ties between verb and object, which together constitute a phrasal unit, the verb phrase (hereafter VP), are evidenced by tests like VP deletion *(Mark washed his car and Bill did [wash the car] too),* VP pronominalization *(Mark washed his car and so did Bill),* and VP fronting *(Mark said that he would wash his car and wash his car he did).* What these tests show is that subjects appear outside the verb phrase in *syntactically unmarked English sentences* (that is, less complex and more frequent than their unmarked counterparts), whereas objects are immediately contained in the VP in English. Subjects and objects, therefore, exhibit different properties of coding and behavior (coding properties are morphological properties such as agreement with the verb, while behavior properties are distributional and include, for example, the omissibility of the subject in a coordinate subject construction like *Matthews complained angrily and left the room*): for instance, in English the verb agrees with the subject—not with the object—in number and person (*He **writes** vs. They **write***) and the subject—not the object—of dependent clauses can be raised to—becomes—subject or object of the matrix (main) clause under certain conditions (*They consider that she is bright/I consider her bright/She is considered bright*).

This correspondence between grammatical functions and syntactic positions or *phrase structure configurations* has motivated the configurational definition of subject and object according to which the subject is the noun phrase directly dominated by the node sentence and the object is the noun phrase directly dominated by the node VP (subconstituents are immediately dominated by the constituents where they belong). This configurational definition puts aside functional relations like subject and object to the benefit of structural aspects of constituency, including, at least dependence (constituents have heads on which subconstituents depend), hierarchy (constituents consist of compulsory and optional subconstituents), and linearization (constituents take up a clausal position relative to the verb).

In this approach, case assignment (as in *who–whom–whose–to whom*) and agreement (as in *I eat* vs. *she eats*) crucially depend on the asymmetry between the noun phrases that bear subject and object: the object requires syntactic continuity with respect to the verb; the subject does not. Syntactic continuity is defined as a relation of adjacency between a head and one of its dependents or between two heads: in *Peter sent them **to the Greens,*** the preposition and the noun phrase it governs are adjacent to each other in the linear order of the sentence, whereas in ***Whom** did Peter send them **to?*** the sequence noun phrase-stranded (that is, final) preposition is interrupted by the subject, the verb, and the object, thus constituting an instance of syntactic discontinuity. Syntactic continuity is often referred to in the literature as *local dependency,* as opposed to *nonlocal dependency* or syntactic discontinuity.

Theoretically speaking, the problem arises of how to account for syntactic discontinuity in grammatical phenomena like case assignment and agreement. In general, two solutions have been adopted: *monostratal* theories recognize a single *level of syntactic representation,* or, more clearly, a single linguistic representation, in which the order and form of the elements of the expansion (including grammatical case and phonological properties of accent, rhythm, and intonation) do not necessarily reflect the order and form of the elements of the sentence that the speaker utters, whereas *multistratal* theories recognize more than one level of syntactic representation, or, in other words, more than one linguistic representation, in such a way that every level represents the sentence that the speaker utters in a less abstract way than the preceding level (abstractness is understood as the distance between the linguistic expression and its representation). Multistratal theories justify the necessity of defining multiple levels of representation on the grounds of the existence of *syntactically marked constructions* (that is, more complex and less frequent than their unmarked counterparts) like passives, questions, preposition stranding, etc., which result from the application of a number of structure-changing operation rules or *transformations* between the deep (more abstract) and the surface (less abstract) levels of syntactic representation. Structure-changing operations fall into three types: first, operations of deletion of specified elements, as in *I met a man **who was wearing a yellow coat*** vs. *I met a **man wearing a yellow coat;*** second, operations of substitution of one specified element by another specified element, as *in These are **the shoes** [I bought the which in Paris]* vs. *These are **the shoes which I bought in Paris;*** and third, operations of permutations of specified elements, as in *Jim doesn´t like **bananas*** vs. ***Bananas** Jim doesn´t like.*

The existence of the node VP, which is criterial for configurationality, has been questioned on a cross-linguistic basis. The interlinguistic evidence provided includes: first, the lack of adjacency between the object and the verb in VSO and OSV languages; second, the syntactically unconstrained permutability of the subject and object in languages such as Latin, where the order of the major elements of the clause can be SVO as well as OVS; and, third, the optionality of overt subject and object in languages like Navajo, a language in which some clauses consist simply of the element V. In consequence, a division into configurational (with VP) and nonconfigurational (without a VP) languages is proposed. Although nonconfigurational languages often coincide with relatively free constituent order languages, this is not always

the case: there are languages, like Navajo, with a relatively rigid constituent order that qualify as nonconfigurational. Navajo has a rich system of verbal prefixes, including prefixes of person and number of the subject and object, thus allowing for *null nominals* (*phonologically empty*, or nonexistent) in the function of both subject and object.

Another dimension has been explored in which languages can be configurational or nonconfigurational: the discoursive dimension. A language is discourse configurational if it assigns the discourse functions of topic and focus. The discourse function topic is used for foregrounding a discourse topic about something that is predicated. Such a discourse topic may or may not coincide with the grammatical subject and is usually placed in a particular structural position, typically clause-initial. The discourse function of focus is used to highlight particular aspects of the predication about the topic. The constituent that bears focus either appears in a particular structural position, typically clause-final, or receives special morphosyntactic coding or prominence, or participates in marked syntactic constructions.

References

Baker, Mark C. 2001. The natures of nonconfigurationality. The handbook of contemporary syntactic theory, ed. by Mark Baltin and Chris Collins. Oxford: Blackwell.

Dik, Simon C. 1989. The theory of functional grammar. Part 1: the structure of the clause, ed. by Kees Hengeveled. Berlin: Mouton de Gruyter, 2nd edition, 1997.

Dik, Simon C. 1997. The theory of functional grammar. Part 2: complex and derived constructions, ed. by Kees Hengeveld Berlin: Mouton de Gruyter.

Hale, Ken. 1983. Walpiri and the grammar of non-configurational languages. Natural Language and Linguistic Theory 1.

Hale, Ken. 1989. On nonconfigurational structures. Configurationality. The typology of asymmetries, ed. by László Marácz and Pieter Muysken. Dordrecht: Foris.

Kiss, Katherine E. (ed.) 1995. Discourse configurational languages, Oxford: Oxford University Press.

Siewierska, Anna. 1988. Word order rules. London: Croom Helm.

Van Valin Jr., Robert D. 2001. An introduction to syntax. Cambridge: Cambridge University Press.

Van Valin Jr., Robert D., and Randy LaPolla. 1997. Syntax: structure, meaning and function. Cambridge: Cambridge University Press.

JAVIER MARTÍN ARISTA

See also **English; Navajo and Athabaskan-Eyak Languages; Syntax**

Connectionism

In the 1980s, the connectionist view, an alternative to rule-based approaches to language, gained increasing acceptance among cognitive scientists. According to this view, our knowledge of language is the result of our interaction with the universe of linguistic utterances that we encounter.

The principal tool of connectionist research to study the acquisition and use of linguistic knowledge is the neural network. A neural network is a computer simulation inspired by neural processes that govern the brain. Neural networks consist of primitive nerve-like processing units, connected at multiple points. The strength of these connections depends on past experience. Accumulated experiences shape the connections, using weights that represent the knowledge of the network. Environmental events are registered via input units as patterns of activation, where the activation is expressed numerically. Thus, the vector (0, 1) indicates activations of 0 and 1 for two input units.

Networks vary in complexity in the number of layers of units and patterns of connectivity. A unit transforms input signals into an activation function and passes the activation forward to other units; therefore, information in neural networks is represented by patterns of activation spread over many neurons and the links between them. The networks are said to represent knowledge in a distributed manner, as a single network can accommodate multiple pairs of independent input–output patterns.

Learning is a natural property of connectionist networks. Learning involves a change of the connection weights in the network as a result of a specific training regimen. Connectionist networks are applied to a wide variety of linguistic phenomena, including recognition, pronunciation, morphology, and syntax, typically with an eye toward simulating empirical effects observed in psychological research. The interactive nature and sensitivity to multiple inputs of connectionist networks are well suited to simulating the recognition of various linguistic structures. An early interactive model by McClelland and Rumelhart (1986) was capable of simulating the word superiority effect, the finding that letters are recognized better when they occur in the context of words rather than in isolation.

Connectionist researchers have also trained networks to acquire spelling-to-sound correspondences by associating orthographic input signals with corresponding phonological outputs. Networks have been trained to mimic the acquisition of morphological knowledge of speakers, for example, of the regular and irregular past tense of English verbs, indicating that children more easily learn those irregular verbs that use the stem for the past tense, as in *hit*.

At least two classes of models of syntactic parsing, or sentence construction are advanced through connectionist modeling. One interprets sentences via case role assignments, while the other constructs phrase structure trees. In both approaches, multiple constraints change as a listener progresses through a sentence. In the first model, each successive word in a sentence imposes constraints on the interpretation of the current case assignment and the entire sentence. The second approach uses a recurrent network to acquire and to discover the phrase tree structure of sentences.

Although increasingly popular for these and other applications (e.g. the lexicon, language impairment, and language production), neural networks remain controversial. According to Pinker (1999), the models suffer from an inherent design flaw, as they do not take advantage of the computational device of a variable, which can stand for a class of words regardless of the particulars of context. The connectionist models are limited by their contradiction between the averaging mechanism and the value of information about the individual nuances so common in language.

Finally, although networks acquire their knowledge without programmer intervention, the programmer has great latitude in shaping the architecture and the processing details of a network. Frequently, in a manner not apparent to others, a theorist determines all of the features of a model, including the number of input units, hidden units, and output units, as well as the pattern of connections between units. While the networks can simulate many empirical findings, there have been both failures of prediction and very powerful predictions, for example, in the domain of speech recognition.

In spite of these criticisms, the neural network approach has become a part of the landscape in cognitive science and—due to its psychological plausibility—will continue to be a significant player in the field.

References

Anderson, James. 1995. An introduction to neural networks. Cambridge and London: MIT Press.

Bechtel, William, and Abrahamsen, Adele. 1994. Connectionism and the mind: an introduction to parallel processing in networks. Oxford (UK) and Cambridge (USA): Blackwell.

Ellis, Rob, and Glyn Humphreys. 1999. Connectionist psychology: a text with readings. Hove (UK): Psychology Press.

McClelland, Jay L., David. E. Rumelhart, and the PDP Research Group (eds.) 1986. Parallel distributed processing: explorations in the microstructure of cognition, Vol. 2: psychological and biological models. Cambridge, MA: MIT Press.

Pinker, Steven. 1999. Words and rules—the ingredients of language. New York: Basic Books.

Rumelhart, David E., Jay L. McClelland, and the PDP Research Group (eds.) 1986. Parallel distributed processing: explorations in the microstructure of cognition, Vol. 1: foundations. Cambridge, MA: MIT Press.

KARL HABERLANDT

Constituency Test

Constituency and Phrase Structure

Constituency has played an important role in grammatical theory, especially regarding the empirical evaluation of different approaches to phrase structure. A constituent corresponds to a unit in the syntactic structure.

Consider the sentence *The janitor read magazines in the morning*. One can account for it as representing a sentence S, containing a noun phrase (NP) *the janitor* and a verb phrase (VP) *read magazines in the morning*.

Both NP and VP are units of S, they are constituents of S. The VP in turn can be analyzed as containing the verb *read*, the NP *magazines* as well as the prepositional phrase (PP) *in the morning*, which are all constituents of the VP. A formal representation in which the major constituents of the sentence are labeled would thus be:

(1) $[_S [_{NP}$ The janitor] $[_{VP}$ read $[_{NP}$ magazines] $[_{PP}$ in the morning]]]

Constituency Tests

Many grammatical operations can only target strings of words that form constituents. This has given rise to a number of widely recognized *Constituency Tests*. The rationale is that a string of words that are taken to form a constituent has to pass one or more constituency tests. The most common tests are summarized below.

Replacement or Substitution Test

A constituent can be replaced by a proform (e.g. a pronominal — *it, he* — or expressions such as *there, then, do so*). The replacing expression can be interpreted in a context so that the meaning of the replaced constituent is completely recovered. For instance, both the underlined subject NP (*a janitor*) and the entire VP in (4) can be replaced by proforms (*he* and *did so*, respectively), showing that they are constituents.

(4) [NP The janitor]$_j$ [VP read magazines in the library]$_k$. [*He*]$_j$ [*did so*]$_k$ everyday.

On the other hand, there is no proform that can substitute for a string of words that spans over parts of different constituents, such as [the janitor read].

Displacement Tests

A constituent can be realized in a position that is different from its canonical position in a sentence, i.e. it appears to be 'displaced'. There are different types of displacement, depending on what additional elements are added to the sentence. In *Clefting*, a structure like (5a) is generated with the addition of *it was…that* supporting dislocation of *in the morning*. However, *magazines in the morning*, which is not a constituent, cannot be clefted, as (5b) shows (the asterisk shows ungrammaticality):

(5) a. *It was* [PP in the morning] *that* the janitor read magazines.
    b. * It was [magazines in the morning] that the janitor read.

In *Preposing,* the material added to the dislocated constituent is usually a form of the copula *be* and a wh-word (e.g. *what, who, where*). The bracketed (embedded) sentence in (6a) is a constituent, and thus it can be displaced, as the preposing structure in (6b) shows:

(6) a. The janitor said [$_S$ that he is on vacation this week].
    b. [$_S$ That he is on vacation this week] *is what* the janitor said.

In *Topicalization,* no material is added to the dislocated constituent. Topicalization is fairly restricted in

English; hence, the embedded clause in (6a) cannot be easily topicalized.

(7) *[That he is on vacation this week] the janitor said.

This restrictive character of topicalization yields an important conclusion. The fact that a given string of words satisfies a constituency test (i.e. any test) is in general sufficient to indicate that this string is a constituent. However, this does not mean that *all* constituency tests can equally felicitously be applied, since each type of sentence structure may be subject to certain constraints (as, e.g. in topicalization).

Coordination Test

The coordination test shows that only constituents can be conjoined (8a), with the additional requirement that they correspond to the same kind of phrase, as shown by the ungrammaticality of (8b):

(8) a. The janitor read magazines [PP in the morning] and [PP in the afternoon].
    b. *Sue drove [NP a car] and [PP to Ottawa].

Relevance of Constituency

Constituency has been a powerful test to evaluate and select among different approaches to phrase structure and grammatical theory. For example, it empirically supported the adoption of X'-Theory in the 1980s, which claimed that the major sentence constituents in turn have internal structure. For instance, *read magazines* in (1) can be shown in (9) to behave as a constituent, excluding *in the morning*.

(9) [Read magazines] *was what* the janitor did in the morning.

This may be taken as evidence for a more fine-grained phrase structure with an additional level between V(erb) and VP, called V' (hence X'-Theory):

(10) [$_{VP}$ [$_{V'}$ read [$_{NP}$ magazines]] [$_{PP}$ in the morning]]]

References

Bunt, Harry, and Arthur van Horck. 1996. Discontinuous constituency. Berlin: Mouton de Gruyter.

Carnie, Andrew 2002. Syntax: a generative introduction. Oxford: Blackwell.

Chametzky, Robert. 2000. Phrase structure: from GB to minimalism. Malden, MA: Blackwell.

Chomsky, Noam 1981. Lectures on government and binding. Dordrecht: Foris.

Hawkins, John A. 1994. A performance theory of order and constituency. Cambridge: Cambridge University Press.

Huck, Geoffrey J., and Almerindo E. Ojeda. 1987. Syntax and semantics, Vol. 20: discontinuous constituency. Orlando: Academic Press.

Kayne, Richard. 1984. Connectedness and binary branching. Dordrecht: Foris.

Lasnik, Howard. 1999. Minimalist analysis. Malden, MA: Blackwell.

Postal, Paul. 1964. Constituency structure: a study of contemporary models of syntactic description. International Journal of American Linguistics 30 (1, part 3).

Siewierska, Anna. (ed.) 1998. Constituent order in the languages of Europe. Berlin: Mouton de Gruyter.

ACRISIO PIRES

See also **Phrase Structure; Pro-forms; Syntactic Category**

Context

Contexts are texts' 'nontexts' and thus relational phenomena. They are caused by figure/ground mechanisms where texts are focused figures and their contexts function as their (back)grounds. Contexts are accordingly always changing. Particular academic fields prioritize particular kinds of researched objects, and accordingly particular kinds of contexts. An overall workable definition of context consequently may not be found, since it would be too general to cover the diverse and specific needs of subdisciplines. This tension between specific and general is captured in the aphorism *Meaning is context bound, but context is boundless* (Culler 1981:24).

Nevertheless, contexts as phenomena have become increasingly important within linguistics, and in general. Everyday configurations and common-sense understandings, reflected in different dictionary definitions, are commonly twofold. Contexts are seen as textual elements embedding particular utterances, and as circumstances surrounding particular situational events. Yet, new insights into how language functions challenge these definitions. First, the notion of text is expanding, incorporating even the meanings associated with pictures, films, and bodies. Thus, an extended concept of text destabilizes a seemingly safe ground for a definition of context. Second, contexts are seen as interactive and not just as passive 'embedding' phenomena, a position that triggers new questions.

When is a context? To report observed situations implies both decontextualizing and recontextualizing. Through decontextualization, focused events are separated from their original environment, and hence their original contexts are destroyed. In order to reconstruct meaning for receivers of reports that are separated from the original context, a text needs to be sufficiently (re)contextualized. However, the report now consists of elements of both the 'original' text and elements of its

new context. Hence, texts, traveling through time, space, and culture, will always occur in new contexts. The *when* question thus leads to a *where* question.

A simple answer to *Where is context?* is: internal and/or external. An internal perception implies seeing contexts as general premises for interpreting utterances, while an external perception perceives the material world as constitutive for contexts. However, both views tend to be static. A dynamic view implies that contexts are not just *around* texts as pre-fixed situations, but are dynamically woven *into* utterances while uttering. Hence, there are contextual elements or hints *within* most texts, which assist the interpretation of the intended meaning. The merging of given (G) and new (N) in texts is regulated by such hints: *Once upon a time there was a boy* [boy N]. *The boy* [boy G] *had a nut* [nut N]. *The nut* [nut G] *had a little hole* [little hole N], etc.

However, the opening phrase, *Once upon a time...,* only seems to imply a fairy tale. Rather, it is an *example* of just another context, namely the genre and context of an entry in the genre and context of an encyclopedia. The outcome is an inclusive embeddedness characteristic for contexts as phenomena [encyclopedia [entry [example [fairy tale hint]]]]. Hence, utterances are *doubly contextual,* relying on existing contexts for their production and interpretation and being, in their own right, events that shape new contexts for following actions.

How 'given' and 'new' are combined depends on the kinds of communication, such as discourses, registers, or genres. Concrete texts and utterances focus on the micro level, while discourses, registers, and genres are active as contexts at a macro level. They can create or are presuppositions for contexts of situations.

Some text theories try to specify contextual constituents more systematically, as M.A.K. Halliday and

his followers (e.g. C.M.I.M. Matthiessen and Jim Martin) do in Systemic Functional Grammar. This approach is contextual and sees language principally as meaning-potential for combining text and context and suggests that language has three interactive main functions: textual, ideational, and interpersonal. Similarly, Jürgen Habermas (1984) sees each person's communicative context as a 'lifeworld' consisting of a triad of three basic interactive elements: inner nature (self), outer nature (world), and others (society). Although he is not arguing that lifeworld *is* context, his triad can be linked to Halliday's so that the textual function is respectively related to communicators, the ideational to world, and the interpersonal to society, so that all elements define each other mutually.

Following Halliday and Habermas and others on this matter, a triadic understanding will imply that both language (or utterances and texts) and context (genres and discourses) will be triadic. They are all interrelated and constitute a systemic *contextual* lifeworld. Consequently, contexts are always and simultaneously embodied, world-related, and societal.

Embodied contexts, however, will in turn function as discursive resources for uttering and as such distribute power asymmetrically, even in a context that is seemingly equal for all the participants. Thus, classroom episodes with students and teacher, medical consultations between patients and doctor, command situations between soldiers and officer, will represent not only different embodied communicational resources, but different contextual power relationships.

If contexts are systemic and interrelated, academic fields have to find their place in such a system, by developing awareness of similarities and differences vis-à-vis other fields. Subdisciplines within linguistics are likely to focus on textually established contexts. The nearest contexts for sounds are words, the nearest contexts for words are sentences, etc. Fields such as applied linguistics and anthropological linguistics, for instance, will need to take the larger, more general contextual aspects, such as genres and discourses, into consideration as relevant for the (re)production of meaning.

Context can be seen as a relational concept in transition and hence open to reconceptualization. To rethink something means to recontextualize it. Thus, reconceptualization is recontextualization.

References

Bakhtin, Mikhail. 1986. Speech genres and other late essays. Austin: University of Texas, Press.

Bateson, Gregory. 1972. Steps to an ecology of mind. New York: Ballantine Books.

Bourdieu, Pierre. 1989. Outline of a theory of practice. Cambridge: Cambridge University Press.

Culler, Jonathan. 1981. The pursuit of signs: semiotics, literature, deconstruction. Ithaca: Cornell University Press.

Duranti, Alessandro, and Charles Goodwin (eds.) 1992. Rethinking context. Language as an interactive phenomenon. Cambridge (UK): Cambridge University Press.

Erickson, Fredric, and Jeffrey Schultz. 1981. When is a context? Ethnography and language in educational settings, ed. by Judith Green and C. Wallat. Norwood, NJ: Ablex.

Fairclough, Norman. 1992. Discourse and social change. Cambridge (UK): Polity Press.

Foucault, Michel. 1972. The archeology of knowledge. London: Tavistock.

Goodwin, Charles, and Alessandro Duranti. 1992. Rethinking context: an introduction. Rethinking context. Language as an interactive phenomenon, ed. by Alessandro Duranti and Charles Goodwin. Cambridge (UK): Cambridge University Press.

Habermas, Jürgen. 1984. The theory of communicative action, Vol 1. Boston: Beacon.

Halliday, M.A.K. 1978. Language as social semiotic. London: Arnold.

Halliday, M.A.K. 1994. An introduction to functional grammar, 2nd edition. London: Arnold.

Malinowski, Bronislav. 1935. Coral garden and their magic, 2 vols. London: Allen & Unwin.

Martin, Jim. 1997. Analysing genre: functional parameters. Genre and institutions. Social processes in the workplace and school, ed. by Frances Christie and Jim Martin. London and Washington: Cassell.

Matthiessen, C.M.I.M. 1993. Register in the round: diversity in a unified theory of register analysis. Register analysis: theory and practice, ed. by Mohsen Ghadessy. London: Pinter.

SIGMUND ONGSTAD

See also **Halliday, Michael Alexander Kirkwood**

Conversation Analysis

Conversation analysis (CA) involves the study of spoken interactions. In CA, naturally occurring talk is recorded, transcribed, and then analyzed. The term conversation analysis, however, may be misleading, as a range of spoken interactions are studied under the heading 'conversation', from casual conversation to

more formal talk such as courtroom interactions or medical consultations. As a result, CA is increasingly referred to as the study of 'talk-in-interaction', although this term has also been criticized for ignoring nonverbal issues such as pauses, gestures, or gaze.

Traditionally associated with sociologist and others working in the ethnomethodological tradition, CA is used in different ways by different scholars. Scholars from a variety of fields, such as linguists, anthropologists, and communication specialists, also analyze spoken interactions, using a range of methods, and based on varying assumptions.

Conversation analysis began in the 1960s with the work of American sociologists Harvey Sacks, Emanuel Schegloff, Gail Jefferson, and others, influenced by Erving Goffman's focus on everyday interactions and by Harold Garfinkel, the founder of ethnomethodology. Ethnomethodology is concerned with uncovering participants' (as vs. researchers') perspectives on the activities in which they are involved. Ethnomethodologists seek to understand how people produce and understand daily activities, including conversation.

Within this tradition, CA sees talk as social action, and attempts to uncover the structural properties of talk in terms of the sequential development of talk. CA provides a substantive body of evidence indicating that talk is rule-oriented. One area of investigation is turn-taking. CA revealed that speakers tend to obey a rule of one-speaker-at-a-time. Several features of talk are related to this preference, including when speakers chose to begin their turns. Another finding is the adjacency pair, a two-turn sequence of adjacent utterances by successive speakers. The first utterance sets up an expectation for the second. An example is greeting–greeting. When a speaker greets another speaker, there is an expectation that the next utterance will be a return greeting. Other areas of investigation in CA have included the concept of repair, or what happens when a speaker misspeaks or is not understood, laughter, and storytelling.

A number of assumptions underlie this methodology used within CA, the ideas that data should consist of multiple instances of naturally occurring talk, not constructed examples, that talk should be transcribed in such a manner as to attempt to capture the sense of the sequential development of the talk, of how the interaction unfolded, and that no item is too small to be investigated. As a result, transcriptions include such items as overlapping speech of multiple speakers talking at once, pauses, laughter, and changes in volume. In addition, CA tends to dismiss assumed categories of sociocultural context such as status or gender, claiming that the relevance of these categories must be demonstrated within the data analyzed.

The term 'context' is used to refer to the textual environment, rather than the sociopolitical environment. Talk can be both context bound, in that we understand utterances because of their sequential placement within a conversation, and context free. Conversationalists seek to find meaning, interpreting utterances in relation to prior utterances or knowledge, and in this sense, all utterances are context bound. Conversation is context free in that there are rules of conversation that are general and are not constrained by changes in such things as the participants or setting: participants in talk still take turns, repair misunderstandings, and the like.

Context is also both retrospective and prospective. What is said in a given moment emerges in part because of what was said in immediately prior utterances, and it also influences what will be said next. In this manner, participants mutually construct talk. CA within the ethnomethodological tradition has a number of limitations. Much of the work in the field has been carried out in English. Cross-cultural studies have not always concurred with its findings. CA has also been criticized for ignoring the larger influences of sociocultural context, and for not acknowledging the benefits of quantitative analyses of spoken interactions.

Scholars from a variety of fields have undertaken analyses of spoken interactions, often addressing the limitations mentioned above. Dell Hymes's ethnography of speaking, also known as the ethnography of communication, is one such area that explicitly examines larger sociocultural contexts such as participant roles or setting. One goal of this approach is cross-cultural comparisons of interactions. Work on speech acts focuses on talk as performing social actions such as complimenting or apologizing. Analysts in the Gricean tradition also tend to use constructed examples of talk in order to determine the cooperative principles of conversation.

Interactional sociolinguistics provides another avenue into the analysis of spoken interactions, as does variation analysis. Much of this work examines how people vary their talk as a result of changes in sociocultural contexts.

Conversation analysis has contributed to linguistics, in part by making talk itself a focus of serious academic investigation. We now understand a great deal about the structure of talk, and therefore how deviations from our expectations of how talk should proceed lead us to interpret—rightly or wrongly—our speech partners' intentions. Fail to answer a greeting with a greeting, for example, and you may well be judged as rude in certain speech communities. CA is also combined with other linguistic disciplines and approaches, such as second-language acquisition and syntactic analysis, investigating how syntactic choices

are affected by the emerging nature of discourse. As CA is incorporated into wider circles of linguistic research, a more complete understanding of the various components and functions of language is possible.

References

Drew, P., and J. Heritage (eds.) 1992. Talk at work: interaction in institutional settings. Cambridge: Cambridge University Press.

Eggins, S., and D. Slade. 1997. Analysing casual conversation. London: Cassell.

Hutchby, I., and R. Wooffitt. 1998. Conversation analysis: principles, practices, and applications. Oxford: Polity Press.

Markee, N. 2000. Conversation analysis. Malwah, NJ: Lawrence Erlbaum.

Psathas, G. 1995. Conversation analysis: the study of talk-in-interaction. Thousand Oaks, CA: Sage.

Sacks, H., E. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organization of turn-taking in conversation. Language 50.

Tannen, D. 1984. Conversational style: analyzing talk among friends. Norwood, NJ: Ablex.

ten Have, P. 1999. Doing conversation analysis: a practical guide. London: Sage.

SUZANNE SCOTT

See also **Context; Discourse Analysis; Hymes, Dell H.; Sociolinguistics; Speech Acts**

Coordination

A coordination joins two sentence elements, called conjuncts. In a coordinate structure like *cats and dogs,* the conjunction coordinates the conjunct *cats* with the conjunct *dogs*. In many languages, conjunctions like *and* or *or* can conjoin words or phrases of virtually every category, under the condition that the categories being conjoined are of the same sort. It might appear as if coordination was a relatively simple phenomenon. However, coordination is notoriously difficult for linguistic theory to define.

Although a wide variety of structures can be conjoined, not all coordinations are acceptable. One of the first generalizations regarding coordination is Ross's Coordinate Structure Constraint (1967). This constraint states that coordination does not allow for asymmetrical constructions. For example, the sentence *This is the man whom Kim likes and Sandy hates Pat* is unacceptable, because only the first conjunct is relativized. The sentence *This is the man whom Kim likes and Sandy hates* is acceptable, because both conjuncts are relativized.

The Coordinate Structure Constraint might be explained by the requirement that the conjuncts in a coordinate construction must be of the same 'sort'. This requirement is sometimes referred to as the Law of Coordination of Likes. Linguists are uncertain as to the relationship between 'sort' and syntactic category. The sentence *Pat is stupid and a liar* shows that being of the same syntactic category is too strong a requirement for conjuncts in a coordinate construction, since an adjective phrase (*stupid*) can be conjoined with a noun phrase (*a liar*). It is therefore unclear what it means for two conjuncts to be of the same sort.

Linguists are further concerned with which material is allowed as a conjunct in a coordinate construction. The second example showed conjoined sentences, but coordination is also possible for noun phrases as in *the apples and the pears*, verb phrases like *run fast or jump high* and adjectival phrases such as *rich and very famous,* etc. Both sentences and phrases intuitively form meaningful units within a sentence, called constituents; however, not all sentence elements can be constituents. Subject and verb do not form a constituent in some frameworks of generative grammar. However, they can occur together as a conjunct in the sentence *Kim bought, and Sandy sold, three paintings yesterday*. The possibility of this 'nonconstituent coordination' has led a number of linguists to relax the notion of constituency. In categorial grammar, for example, subject and verb can form a constituent. Coordination phenomena can therefore provide a testing ground for even basic theoretical notions such as constituency.

Another important question concerns the way coordination of phrases is interpreted, as phrasal coordination seems strongly related to sentential coordination. The sentence *Kim ran and jumped*, in which two verb phrases are conjoined, has the same interpretation as the coordination *Kim ran and Kim jumped*, in which two sentences are conjoined. The dominant approach in generative syntactic theories indicates that phrasal coordination can be derived from sentential coordination by means of reduction rules. This approach states that *Kim ran and jumped* is the result of a reduction rule having deleted the subject of the second conjunct, *Kim*. Much syntactic research focuses on formulating appropriate reduction rules. This has turned out to be

quite difficult due to the potential for phrasal constructions such as *Kim and Sandy are similar,* which lack a sentential source.

The work of Richard Montague in the early 1970s ushered in the method of deriving the interpretation of conjoined phrases directly from their surface form. Semantic explanations of phrasal coordination differ from strictly syntactic ones. In semantic analysis, conjuncts are interpreted as functions that require certain semantic arguments to make a sentence. The semantic type of *ran* requires a subject to yield an interpretable sentence. Because *jumped* is of the same semantic type as *ran*, *ran* and *jumped* can be conjoined according to the Law of Coordination of Likes. If this conjoined verb phrase is applied to the subject *Kim*, the resulting interpretation will be that Kim ran and Kim jumped. Under this semantic approach, the interpretation of phrasal coordination is related to, but not derived from, the interpretation of sentential coordination.

Further, a problem particular to syntactic analysis concerns how to formally represent coordination, as the conjunction structure appears to contradict current models of sentence construction. The term 'coordination' implies that the conjuncts are located at the same structural level. However, this view is incompatible with the assumption in generative syntax that syntactic structures are formed by repeatedly unifying two elements at a time. The argument that a conjunction unifies its two conjuncts at the same level entails the conclusion that syntactic structures can unify three elements. Some linguists attempt to remedy this apparent contradiction by suggesting a subordinating structure for coordination, in which the conjunction combines with one conjunct first and the resulting constituent then combines with the other conjunct. Others suggest the hypothesis that both conjuncts basically stand in

exactly the same relationship to the conjunction, and that the two conjuncts essentially introduce a third structural dimension.

Although various solutions have been proposed, there is not yet a satisfactory explanation for all of the problems discussed here. Coordination occurs both in phrase structure and sentence structure, but the relationship between the two remains unclear. The role of coordination in sentence construction is also largely undetermined. Coordination therefore remains a phenomenon that is difficult to explain for any formal linguistic theory.

References

Goodall, Grant. 1987. Parallel structures in syntax. Cambridge: Cambridge University Press.

Hartmann, Katharina. 2000. Right node raising and gapping: interface conditions on prosodic deletion. Philadelphia/Amsterdam: John Benjamins.

Johannessen, Janne Bondi. 1998. Coordination, New York: Oxford University Press.

Montague, Richard. 1973. The proper treatment of quantification in ordinary English. Approaches to natural language: Proceedings of the 1970 Stanford Workshop on Grammar and Semantics, ed. by Jaakko Hintikka, J.M.E. Moravcsik and Patrick Suppes. Dordrecht: Reidel.

Oirsouw, Robert R. van. 1987. The syntax of coordination. London: Croom Helm.

Ross, John Robert. 1967. Constraints on variables in syntax. Ph.D. Dissertation, MIT, reprinted as Infinite syntax; Norwood, NJ: Ablex Publishing Corporation, 1986.

Steedman, Mark. 2000. The syntactic process. Cambridge, MA: MIT Press.

Winter, Yoad. 1998. Flexible Boolean semantics: coordination, plurality and scope in natural language. Ph.D. Dissertation, Utrecht University.

PETRA HENDRIKS

See also **Constituency Test; Montague, Richard**

Coptic Egyptian

In the course of its productive language history, which spans a period of over 4,000 years, Ancient Egyptian went through several developmental stages. Its latest stage is Coptic Egyptian (not to be confused with Egyptian Arabic), which is the vernacular of late-antique and medieval Christian Egypt (fourth to fourteenth centuries AD). The modern term Coptic is derived from Arabic *qubṭī*, itself a corruption of the Greek word (*ai*)*gypt*(*ios*) 'Egyptian'.

The Copto-Greek alphabet

Coptic, like many other ancient languages of literature, has been passed down to us through large corpora of texts. These texts were written in a highly standardized notational system of alphabetic signs representing the different sounds of the Coptic language. The origin of the Coptic writing system lies in occasional Greek transcriptions of native words in Egyptian texts of the Hellenistic and Roman periods. In the first three

centuries AD, the use of such transcriptions became increasingly common and entire corpora of texts with a predominantly magical character were written in a Greek-derived alphabet. The Christianization of the country in the fourth century AD constituted a turning point: the abandonment of the pagan literary tradition and culture manifested itself in the replacement of hieroglyphic writing and Demotic, its cursive variant, by Greek script. In its present form, the Copto-Greek alphabet consists of 32 letters, 24 of which are taken from Greek and eight from Demotic writing. With the exception of the monosyllabic grapheme *ti*, the Demotic-based signs represent phonemes that were absent in Greek, but which are part of the native Coptic Egyptian sound system.

Bilingualism and linguistic borrowing

The emergence of Coptic was the result of intensive language contact in a bilingual (Egyptian–Greek) speech community. Greek was not only the language of the literate elite but also the language of the Holy Scriptures and the new religion, and hence a language of great cultural importance. The impact of this prestigious language on the native vernacular was all-pervasive. Although no clear statistics are available at present, it is estimated that approximately 40% of the Coptic vocabulary consists of Greek loanwords. The transfer of Greek lexical material into the Coptic vocabulary was not only restricted to content words (nouns, verbs, adjectives) with a clear link to Hellenistic and Christian culture (e.g. *te-psikhɛ* 'the soul', *klɛronomei̯* 'to inherit', *hagios* 'holy'), but also involved a variety of Greek function words (i.e. grammatical words with no descriptive-lexical content), such as sentence conjunctions, discourse markers, and some types of adverbs and prepositions (e.g. *hɔste* 'such that', *ɛ* 'or', *oude* 'and not', *alla* 'but'). Despite the massive influx of Greek items, grammatically organized words (pronouns, articles, numbers, and the like) are all drawn from the native stock. To fit into Coptic phrase structure, Greek loanwords underwent minor changes in the course of borrowing.

Dialect variation

Coptic Egyptian is actually a dialect cluster, consisting of at least six regional varieties, two of which gained supraregional importance: Sahidic, the language of the whole Nile valley above the Delta, and Bohairic, the language of the Nile Delta. Sahidic and Bohairic Coptic differ significantly from one another in a number of grammatical features (sound system, verbal morphology, syntax) and the amount of lexical and grammatical borrowing from Greek. Both language varieties also differ with respect to the time depth of their attestation. Prior to the Arabic conquest in AD 641, Sahidic was the predominant literary dialect of Coptic. Its supremacy became challenged by Bohairic Coptic from the ninth century onward. By the eleventh century, Bohairic had replaced Sahidic as the official church language and had become the sole representative of Coptic Egyptian, which survived as the liturgical language of the present-day Coptic Orthodox Church. The language material of the following typological sketch is exclusively drawn from Sahidic Coptic, the main reference dialect.

Phonology

The sound system of Sahidic Coptic consists of 20 consonants (p, t, k, d, g, kʲ, f, s, ʃ, h, β, z, tʃ, m, n, l, r, w, y, ʔ) and eight vowels (i, e, ɛ, u, o, ɔ, a, ə). Its most striking property is the absence of voiced consonants (i.e. consonants produced with a vibration of the vocal cords) in the class of stops (consonants produced by making a closure at the lips), fricatives (consonants where the airflow is constricted to form a turbulence but is not interrupted), and affricatives (sounds produced by an initial closure that is released gradually, making it sound like a fricative), the main exception being the voiced fricative /β/. It is worthwhile pointing out that the voiced stops /d/ and /g/ as well the voiced fricative /z/ have a special status as loan phonemes that are by and large restricted to Greek borrowings. The glottal stop /ʔ/ (a sound produced by a complete but brief contraction of the vocal cords) has no alphabetic sign of its own. Yet, its presence in the Sahidic sound system can be deduced from a sequence of two identical vowel letters that indicates 'broken' vowels, i.e. long vowels that are pronounced with a brief interruption (e.g. *maatʃe* /maʔatʃe/ 'ear', *mɛɛʃe* /mɛʔɛʃe/ 'crowd'). Coptic has pairs of vowels: /e/ vs. /ɛ/ and /o/ vs. /ɔ/. The schwa /ə/ (a colorless, unstressed vowel) is graphically expressed not by a letter, but rather a special diacritic, the so-called supralinear stroke (a vertical line above a letter), e.g. *n̄* /ən/ 'of'. Stress (indicated as ˈ) is lodged only on those syllables that either contain a long vowel (ː) (e.g. *noute* /nuːˈ.te/ 'god') or a vowel–consonant sequence (e.g. *anaʃ* /a.naˈʃ/ 'oath'). Moreover, there is at most one main stress per word, regardless of its length.

Noun morphology

Unlike Pre-Coptic Egyptian, Coptic has only a few word formation processes for the makeup of nouns. Thus, there are only a few lexical items where number (singular vs. plural) and gender (masculine vs. feminine) are marked on the noun itself (e.g. *son* 'brother',

sɔne 'sister', *sneṷ* 'brothers, fellow monks'). In the vast majority of cases, gender and number distinctions are expressed on the definite article, which is placed in front of the noun it modifies (e.g. *p-rɔme* 'the (sing. masc.) man', *te-shime* 'the (sing. fem.) woman', *ne-shime* 'the (pl.) women'). The Sahidic determiner system offers a three-way contrast between a definite, an indefinite, and a zero article (e.g. *p-rɔme* 'the (sing. masc.) man'; *u-rɔme* 'a man', *hen-rɔme* 'men'; *rɔme* 'man'). Besides the indefinite and definite articles, there are demonstratives articles, so-called because they indicate closeness or distance to the location of the speaker: *peḭ-rɔme* 'this (sing. masc.) man' vs. 'that man' (lit. the man who is (over) *there*). Both demonstrative articles have corresponding demonstrative pronouns, which have a word-like status: *paḭ* 'this one' vs. *p-et-əmmaṷ* 'that one'.

Reference to grammatical person (the speaker, the addressee, and a third party) is expressed by two sets of pronouns: independent pronouns and bound pronouns. Independent pronouns have a word-like status, while bound pronouns form an inseparable unit with the sentence element they modify. Different forms exist for pronominal prefixes that precede and pronominal suffixes that follow their host. Table 1 presents the pronominal paradigm of Sahidic Coptic (-∅ indicates a null morpheme, i.e. a grammatical element that is not pronounced). Independent pronouns have an emphatic meaning and place the entity or object they refer to at the center of attention. Pronominal prefixes and suffixes, on the other hand, lack such a highlighting function. While independent pronouns are restricted to peripheral syntactic positions (e.g. *ntos de a-s-onk-s ehraḭ* '(as for) her (ntos), she leaped up herself' (Eudoxia 50:17)), bound pronouns are compatible with all nominal positions of the clause (e.g. *a-f-ent-s* 'he brought it', *nɛ-tən* 'for you').

Verb morphology

The basic principle of Coptic verb formation, which it shares with Semitic languages, is that of 'root-and-pat-tern'. As the terminology suggests, verbal stems are derived from relatively abstract form-meaning representations (roots) by the superimposition of particular consonant–vowel patterns with a basic meaning. For example, a verbal root like /*m-s* / 'BEAR, GENERATE' may surface in at least four word formation patterns that are distinguished from one another by means of vowel change (Ablaut) and syllable structure. These are the absolute state *mise* 'to give birth', the nominal state *mes,* and the corresponding pronominal state *mest* 'to deliver', which combine with a nominal and pronominal object, respectively, and the Stative *mose* 'to be bred'. The absolute state and the construct state (i.e. the nominal and the pronominal states together) describe events that *change* over time, while Statives refer to states or conditions that *last* for some time. The absolute state and the construct states, in turn, differ syntactically in the way they express the direct object relation. In the absolute state, the direct object is encoded as a prepositional phrase (e.g. *mise n-u-ʃeʔere n-shime* 'to give birth to a girl' (Miracles of Apa Mêna 10[b]:33–4)). In the construct state, however, direct objecthood is indicated by the juxtaposition of the verb and the nominal or pronominal object (e.g. *mes p-ʃr-howt* 'to deliver the male child' (Apocalypse 12:13) vs. *mest-f* 'to deliver it' (Miracles of Apa Mêna 10[b]:26)).

Conjugation system

Coptic has more than 20 different verb conjugations (i.e. the pattern in which verb forms can appear) for the expression of tense (i.e. the location of events in time), aspect (i.e. the ongoing state, completion, or multiple occurrence of events), and mood (i.e. the commitment of the speaker toward the truth of the reported events). At the foundation of this richness of meaning distinctions is the subdivision of the four absolute tenses (the Present, the Habitual, the Future, and the Perfect) into two conjugation classes, traditionally known as first and second tenses. The second tenses are morphologically derived from the 'basic' first tenses by adding the relative markers *e-* or *ənt-* in front of the verbal cluster (Table 2).

First and second have exactly the same temporal and aspectual interpretation, but differ from one another with respect to their syntactic distribution. First

TABLE 1 Pronominal System

	Bound Pronouns		Independent Pronouns
	Prefixes	Suffixes	
1st sing.	ti-	-i, -t	anok
2nd sing. masc.	k-	-k	əntok
2nd sing. fem.	te-, ter-	-e, -∅	ənto
3rd sing. masc.	f-	-f	əntof
3rd sing. fem.	s-	-s	əntos
1st pl.	ten-	-n	anon
2nd pl.	tetən-	-tən, -teṷtən	əntotən
3rd pl.	se-	-se, -suː, -uː	əntoṷ

TABLE 2 Absolute Tense System

First Tenses		Second Tenses	
Present	f-sɔtəm	e-f-sɔtəm	'he listens'
Habitual	ʃa-f-sɔtəm	e-ʃa-f-sɔtəm	'he (usually) listens'
Future	f-na-sɔtəm	e-f-na-sɔtəm	'he is going to listen'
Perfect	a-f-sɔtəm	ənt-a-f-sɔtəm	'he listened'

tenses appear in pragmatically neutral declarative clauses (e.g. *eịs hɛʔɛte anon ne.k-həmhal tən-sɔtəm* 'Look, *we*, your servants, *are listening*!' (Eudoxia 62:03)). The derived second tenses are used in the context of relative clauses (e.g. *u-hɔβ* [*ere p-nuːte moste əmmo-f*]'a thing [that God *hates* (it)]' (Acts of Andrew & Paul 202:126–7), constituent questions (e.g. *e-tetən-ʃine ənsa nim*'Whom *are you* (woman) *looking* for?' (John 18:4), as well as a range of declarative focus contexts (e.g. *awo p-woeịn e-f-r woeịn həm pə-kake* 'and (as for) the LIGHT, *it is shining* in the DARKNESS' (John 1:5)).

Apart from absolute tenses, Coptic has several syntactically dependent verbal tenses and moods. Relative tenses locate some event with respect to another event and express three types of temporal relations: precedence, subsequence, or simultaneity. In doing so, relative tenses often indicate a logical or causal connection between two events. In negative tenses, negative meaning (the equivalent of English *not*) and a particular time value are fused together in a single indivisible unit. Take for instance the Negative Future *ənne-f-sɔtəm* 'he shall not hear', where the conjugation *ənne* combines future tense with negative meaning, which are expressed by two separate elements (viz. *shall* and *not*) in the English equivalent. Moreover, Coptic has two different moods for the imperative (e.g. *sɔtəm* 'Listen!'), which expresses commands, and the optative, which expresses wishes (e.g. *mare-f- sɔtəm* 'may he listen'). See Table 3 for further illustration.

The clause-initial or medial position of tense-aspect-mood markers, their morphological independence from the verb, as well as their agreement behavior (variant forms for nominal and pronominal subjects, e.g. (Habitual) *ʃare p-rome sotəm* 'the man listens (usually)' vs. *ʃa-f-sotəm* 'he listens (usually)'

provides *prima facie* evidence for their categorical status as auxiliary verbs. Due to their semantic erosion, these auxiliaries have a fully grammaticalized meaning and function, which is typical of free functional morphemes.

Syntax

The basic word order, from which other word-order patterns are derived, is subject–verb–object (SVO). Word-order alternations are, however, extremely common and motivated by pragmatic considerations. The topic status (i.e. the presupposed, familiar or known character) of a noun or pronoun is generally indicated by placing it into the left periphery of the clause, its grammatical function being resumed by a pronoun with identical person, number, and gender specification (e.g. *p-aŋgelos de m-p-tʃoeịs a-f-wonh-f e-p-arkhiepiskopos* 'the angel of the Lord, *he* revealed himself to the archbishop' (Mêna, 4b:6–9)). Verb–subject order has a 'presentative' meaning and is used for the introduction of new discourse participants. Notice that the nominal subject that is removed from its sentence-initial position is supplied with the particle *nkʲi* (e.g. *f-nɛu nkʲi u-aŋgelos nte p-nuːte* '(there) comes *an angel* of God' (Budge, Coptic Martyrdoms 214:22)). To highlight the subject or object, a special construction type is used: the cleft sentence. It is called cleft sentence because it consists of two parts: a sentence-initial noun or pronoun and a relative clause (given in brackets) (e.g. *awo m-pe.u-kʲβoi an pe* [*ənt-a-f-tutʃo-ụ*] 'and (it) is not *their arm* that has saved them'(Psalm 43:4)). More research is needed to clarify the relation between pragmatic prominence and sentence form in Coptic Egyptian.

TABLE 3 Relative Tenses and Basic Moods

Relative Tenses		Negative Tenses	
Temporal	əntere-f-sɔtəm 'after he had listened'	Negative Habitual	me-f-sɔtəm 'he does not listen'
Terminative	ʃantə-f-sɔtəm 'until he listens'	Negative Future	ənne-f-sɔtəm 'he shall not listen'
Conjunctive	nə-f-sɔtəm 'and he listens'	Negative Perfect	əmpe-f-sɔtəm 'he did not listen'
Finalis	tare-f-sɔtəm 'so that he will listen'	Unexpected Negative Perfective	əmpate-f-sɔtəm 'he has not yet listened'
		Moods	
Conditional	e-f-ʃan-sotəm 'if he listens'	Negative Conditional	e-f-ʃan-təm-sotəm e-f-təm-sotəm 'if he does not listen'
Optative	mare-f-sɔtəm 'May he listen!'	Prohibitive	mpər-sɔtəm 'Do not listen!'
Imperative	sɔtəm 'Listen!'		

References

Layton, Bentley. 2000. A Coptic grammar with Chrestomathy and Glossary. Porta Linguarum Orientalium N.S. 20. Wiesbaden: Harrassowitz Verlag.

Loprieno, Antonio. 1995. Ancient Egyptian—a linguistic introduction. Cambridge and New York: Cambridge University Press.

Polotsky, Hans Jakob. 1960. The Coptic conjugation system. Orientalia 27. 392–422.

Reintges, Chris H. 2001. Code-mixing strategies in Coptic Egyptian. Lingua Aegyptia 9. 193–237.

Reintges, Chris H. 2003. Coptic Egyptian (Sahidic Dialect)—a learner's grammar. Cologne: Rüdiger Köppe Verlag.

Reintges, Chris H. 2004. Coptic Egyptian (Sahidic Dialect)—a learner's grammar. Cologne: Rüdiger Köppe Verlag.

Till, Walter C. 1961. Koptische Dialektgrammatik: mit Lesestücken und Wörterbuch. München: Verlag C.H. Beck.

CHRIS H. REINTGES

See also **Ancient Egyptian; Auxiliaries; Function Words**

Corpus Linguistics

During the last three decades of the twentieth century, computer technology has made it possible to conduct extensive and complex research on specific linguistic features — either lexical items or grammatical structures—and their systematic associations with other linguistic and nonlinguistic features. These nonlinguistic features include registers or specific varieties of language (e.g. religious, political, scientific) and dialects, which are regional or social varieties of a language. This new type of research is part of *corpus linguistics*, which is the empirical study of language using computer techniques and software to analyze large, carefully selected and compiled databases of naturally occurring language (Conrad 2000).

Corpus linguistics represents a departure from the dominant mentalist approach to linguistic research, which emphasizes the processes taking place in the human mind (e.g. a Universal Grammar-centered approach to second language acquisition studies). This type of research is characterized by the empirical analyses of actual patterns of language use in large and principled databases or corpora. Corpus linguistics utilizes both quantitative and qualitative analytical techniques and relies on computers to perform complex analyses. (See Biber, Conard & Reppen (1998) and Kennedy (1998) for a detailed account of corpus-based investigations of language structure and use.)

Corpus-Based Investigations of Language Use

Although language structure has traditionally been studied using nonempirical methods and relying on the researcher's intuitions, extensive corpus-based studies describing various aspects of language use were carried out in the 1980s and 1990s (Aarts 1991, Aijmer and Altenberg (eds.) 1991, Biber 1995, Leech 1991, Sinclair 1987, 1991, Stubbs 1995, Svartvik 1990).

As Biber et al. (1996:115) point out, 'it is in the area of language use that corpus-based techniques have had the most impact'. These studies complement descriptive language structure investigations and previously neglected aspects of English grammar (Crystal 1979). Early studies in corpus linguistics focused on the occurrence of linguistic items (e.g. noun, verb, and adjective frequencies), but the development of more powerful techniques has enabled researchers to identify and analyze complex association patterns. Corpus-based research has shown that these linguistic association patterns generally fall into two major categories: lexical associations and grammatical associations. In the first case, the goal is to investigate how a linguistic feature is systematically associated with particular words. In the second case, the researcher investigates how a linguistic feature is systematically associated with grammatical features in the immediate context.

With respect to lexical associations, a concordancer shows which words collocate with each target word in a corpus that is representative of a specific register or dialect. For example, in a corpus of working-class northern New Jersey cyber discourse, we find that the mental verb *know,* which appears 15 times in a small sample thread (7,000 total words)**,** collocates primarily with personal pronouns *I, you, they*, and *we.* This usage reflects the informal nature of cyber exchanges, which characterizes casual face-to-face interaction among peers or friends.

The results of empirical, large-scale corpus-based research projects (data from a 5.7-million-word sample from the Longman—Lancaster Corpus) have

---

KWIC Display (TACT)

know (15)

(219) That's how it works and they **know** it. At least now they're ¦

(227) Are you a total moron? Do you **know** the number of lawsuits ¦

(314) and hazardous 33 years ago. I **know** of no environmentall ¦

(329) against Mar...and Cap...You **know** the condition of the ¦

(372) DPW does for this town. I don't **know** what your job is, but ¦

(372) responsibility at all because they **know** you couldn't handle ¦

(403) Just Curious about how…We all **know** that the building is a ¦

(418) he supervised his dog. I don't **know** if your post even ¦

(419) electrician for years, you don't **know** too much about his ¦

(432) you wouldn't want anyone to **know**? Would you? ¦

(460) You are there aren't you? I **know** it's not election ¦

(467) IMPLANT Does Maru… **know** where the D.P.W. ¦

(473) Hard Worker #2: How do you **know** that Caputo visits Kiss ¦

(475) "does Maru…or the mayor **know** where the DPW is?" ¦

(479) to above post: how do you **know** who visits nortons ¦

---

shown that where certain words are nearly synonymous in isolation, careful analysis reveals that they tend to be used with very different kinds of words. For example, Biber et al. (1998, chapter 2) demonstrate that the word *big* commonly co-occurs with *toe,* while *large* commonly co-occurs with *number.*

Linguistic features are also systematically associated with grammatical features in the immediate context as shown by corpus-based research. For example, one of the factors differentiating *that-clauses* and *to-clauses* is their lexical associations. 'The verbs such as *suggest*, *conclude*, *guess*, and **argue** can control a *that-clause* but not a *to-clause*; the verbs *begin*, *start*, and *try* can control a *to-clause* but not a *that-clause*' (Biber et al. 1998, chapter 3).

There are many promising applications of corpus-based research that have to be explored. For example, it is hoped that new grammar teaching materials for English-as-a-Second Language (ESL) learners will incorporate the results of corpus-based research. Conrad (2000) and Granger (2002) explain that corpus research complements innovations in grammar pedagogy by encouraging instructors to design and implement consciousness-raising activities for second and foreign language learners. (See Conrad (2000) for an overview of the applications of Corpus Linguistics in grammar teaching in the twenty-first century, and Granger (1998a; 1998b; 2002) for a discussion of learner corpora compilation processes.)

Corpus Analysis Tools and Corpora Used in Corpus Linguistics

Since corpus linguistics utilizes large and representative collections of natural texts, there are several types of tools that can be used to conduct research: commercially available packages or concordancing programs (e.g. LEXA, MonoConc, MicroConcord, TACT, WordSmith, WordCruncher) and computer programs developed by researchers for specific types of analyses. These latter types of programs are used to investigate complex grammatical constructions or association patterns, such as Biber et al.'s (1998) study involving the omission of *that* from *that-clauses*.

There are at least 31 commercially and publicly available corpora of written and spoken texts. These databases contain millions of words and are divided primarily into three categories: written, spoken, and historical. These corpora are further subdivided into American English, British English, and texts of other varieties of English. Corpora also exist for languages other than English. (See Appendix in Biber et al. 1998: 282–7.)

Potential Limitations of Corpus Linguistics

It is important to understand that meticulous qualitative analyses of single texts (e.g. historical, spoken, written, and learner-centered) are usually undertaken before embarking on corpus-based research. A careful micro-analysis of linguistic features in written or spoken discourse helps us frame new research questions and hypotheses. Thus, corpus-based analysis should be seen as a complementary approach to the more traditional approaches that have often focused on language structure. The strength of corpus linguistics lies in its investigations of language use, which necessitate empirical analysis of large databases of authentic texts.

There are basically four potential limitations of corpus-based research. The first one is the time-consuming nature of compiling a corpus and tagging parts of speech, errors, or other features. The second limitation concerns the nature of the corpus data collected. This type of research often involves collecting and storing large corpora that may need to be continuously updated.

The third limitation is that relying on computers forces linguists to concentrate primarily on written rather than spoken language. Lastly, there is a potential limitation of a philosophical nature, namely that a corpus is a finite sample of an infinite population. This means that researchers are often extrapolating from what is found in a corpus to what is true of the language or language variety it is supposed to represent (Leech 1998). Thus, one needs to be cautious in drawing general inferences from the results of corpus-based analyses.

Despite its potential limitations, almost any aspect of linguistics can be studied from a use perspective. Corpus linguistics provides a variety of tools and methods that make large-scale research on complex linguistic phenomena an extremely productive and challenging undertaking.

References

Aarts, Jan. 1991. Intuition-based and observation-based grammars. English corpus linguistics, ed. by Karin Aijmer and Bengt Altenberg, 44–62. London: Longman.

Aijmer, Karin, and Bengt Altenberg (eds.) 1991. English corpus linguistics. London: Longman.

Biber, Douglas. 1995. Dimensions of register variation: a cross-linguistic comparison. Cambridge: Cambridge University Press.

Biber, Douglas, Susan Conrad, and Randi Reppen. 1996. Corpus-based investigations of language use. Annual Review of Applied Linguistics 16. 115–36.

Biber, Douglas, Susan Conrad, and Randi Reppen. 1998. Corpus linguistics: investigating language structure and use. Cambridge: Cambridge University Press.

Conrad, Susan. 2000. Will corpus linguistics revolutionize grammar teaching in the 21st century? Tesol Quarterly 34. 548–60.

Crystal, David. 1979. Neglected grammatical factors in conversational English. Studies in English linguistics for Randolph Quirk, ed. by S. Greenbaum, Geoffrey Leech and Jan Svartvik, 153–67. London: Longman.

Granger, Sylviane (ed.) 1998a. Learner English on Computer. New York: Addison Wesley Longman Limited.

Granger, Sylviane 1998b. The computer learner corpus: a versatile new source of data for SLA research. In Granger, 3–18.

Granger, Sylviane, Joseph Hung, and Stephanie Petch-Tyson (eds.) 2002. Computer learner corpora, second language acquisition and foreign language teaching. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Kennedy, Graeme. 1999. An introduction to corpus linguistics. London, New York: Longman.

Leech, Geoffrey. 1991. The state of the art in corpus linguistics. English corpus linguistics, ed. by Karin Aijmer and Bengt Altenberg, 8–29. London: Longman.

Leech, Geoffrey. 1998. Preface. In Granger.

Sinclair, J. (ed.) 1987. Looking up. An account of the cobuild project in lexical computing. London: Collins ELT.

Sinclair, John. 1991. Corpus, concordance and collocation. Oxford: Oxford University Press.

Stubbs, Michael. 1996. Collocations and semantic profiles: on the cause of the trouble with quantitative studies. Functions of Language 2. 23–55.

Svartvik, Jan. 1990. The London-Lund corpus of spoken English: description and research. Lund, Sweden: Lund University Press.

TACT Database is under construction (cyber discourse of working-class northern New Jersey towns). Contact Susana Sotillo at Sotillos@mail.montclair.edu.

SUSANA M. SOTILLO

See also **Universal Grammar**

Courtroom Discourse

Studies in courtroom discourse often focus on the significance of certain kinds of interactions between specific linguistic features and their pragmatic or sociointeractional functions in the courtroom. Research carried out in North Carolina beginning in the 1970s, for instance, resulted in a series of publications suggesting broadly that lay witnesses are persuasive or not depending upon the extent to which they use "powerful" or "powerless" language. The validity of the study's methodology has been questioned, but the publications generated by the study have clearly served to advance our understanding of the role of language in judicial process, particularly where the questioning of lay witnesses is concerned.

Other research has explored the questioning tactics used by attorneys during direct and cross-examination of witnesses. The use of open-ended wh-questions (who, what, where, when, how, etc.) tends to elicit narrative responses from witnesses, and therefore more information, whereas yes/no questions tend and are usually intended to limit the amount and type of information that can be offered. A number of studies examine such patterns, and it is usual for trial practice courses in law schools to include material on such topics, often recommending that wh-questions be used during direct examination and that yes/no questions be used during cross-examination. For a recent example of such a study, see Sandra Harris' (2001) examination

of the O.J. Simpson, Louise Woodward, and Oklahoma federal building bombing cases.

It is usual for defense attorneys to omit mention of the agent in cases of sexual assault or rape; the tactic is thought to take the focus off the accused rapist. The grammar and prosody of reported speech can be used to impeach or cast doubt on a witness's testimony. Janet Cotterill examined the O.J. Simpson trial and explored how metaphors can be a rich source of indirect messages, noting that several were used in defense attorney Johnny Cochran's closing arguments.

In a more general sense, courtroom discourse includes attention not only to adversarial questioning techniques, particularly as used in direct and cross-examination, but also to all spoken language used during the trial process. In order to look meaningfully at the full range of courtroom discourse, one must recognize that a number of speech genres are used during the trial process. Anglo-American lawyers like to talk about the need for a theory of the case. Once that theory is developed, all portions of trial process are theoretically dedicated to developing that theory and persuading the judge and/or jury that it is the correct one. In the process of moving through trial process, lawyers make use of opening statements, voir dire, direct examination, cross-examination, possible redirect and recross, and then closing statements. They also usually participate in the wording of jury instructions, and they make and respond to objections during testimony. For each of these steps, discourse strategies are developed.

Based on an examination of a single case, Gail Stygall (1994) suggests that analysis of legal language needs to move away from mere descriptive analysis to an examination of how the maintenance of legal language serves institutional power and dominance.

More specifically, John Gibbons (2003) suggests that the very real power and dominance problems in the courtroom arise from the legal genres used in courtroom discourse and the fact that they are addressed to two very different audiences, lay persons, often jurors, on the one hand, and of course judges and lawyers on the other hand. Gibbons calls this the two-audience dilemma and cites jury instructions as a particularly difficult genre, because "[j]urists persist in administering these instructions to jurors because they have survived on appeal to a higher court... The view that jurors remember and understand any instructions given them and that therefore it is more important to focus on an instruction's survivability on legal appeal than on its comprehensibility, is ill-informed" (English and Sales 1997. 383).

Where jury instructions are concerned, it is an established fact that since the early 1970s, findings by many social scientists confirm that lay persons are frequently bewildered by the wording of jury instructions. The exact syntactic and semantic bars to juror comprehension of instructions are now well documented by linguists and psycholinguists. These scholars have demonstrated that instructions can be made more comprehensible by simplifying sentence structure and by giving additional information about the meanings of abstract terms in both civil and criminal cases.

We know a great deal about the problems faced by lay jurors because of the extensive amount of research that has been completed. But lay witnesses face similar problems. They contend with very sophisticated question-and-answer tactics on direct and cross-examination, usually with little or no preparation. And, of course, court interpreters and translators face the same problems, even as they deal with the challenge of working with more than one language.

All phases of pretrial, trial, and posttrial (appeal) procedure involve judicial procedures in which lay persons must deal with legal language and discourse patterns, including legal genres, with which they are quite unfamiliar. Jury instructions have been examined in great detail; research on other patterns and genres remains to be done.

It seems likely that much additional research will be reported over the next few years. Increasing numbers of linguists have been focusing on language and law issues, and there are now at least two academic organizations that regularly focus on the kinds of topics involved in courtroom discourse. Since 1990, the Law and Society Association, an international association that meets outside the United States every third year, has scheduled sessions on Language and Law at its annual meetings. The International Association of Forensic Linguists (IAFL) came into existence as a result of a series of workshops and conferences in the early 1990s; it meets biannually, usually outside the United States The IAFL publishes a journal, until recently titled *Forensic Linguistics: The International Journal of Speech, Language and the Law*. Future issues will drop the original title and use the original subtitle only.

References

Coates, Linda, Janet Beavin Bavelas, and James Gibson. 1994. Anomalous language in sexual assault trial judgments. Discourse & Society 5(2 April).189–206.

Conley, John M., William O'Barr, and E. Allen Lind. 1978. The power of language: presentational style in the courtroom. Duke Law Journal 78. 1375–99.

Cotterill, Janet. 1998. 'If it doesn't fit, you must acquit': metaphor and the O. J. Simpson criminal trial. Forensic Linguistics: The International Journal of Language and the Law 5(2). 141–58.

Cotterill, Janet. 2002. 'Just one more time…' aspects of intertextuality in the trials of O.J. Simpson." Language in the Legal Process, ed. by Janet Cotterill, Chapter 9, Basingstoke and New York: Palgrave. 147–61.

Dumas, Bethany K. 2000a. Jury trials: lay jurors, pattern jury instructions, and comprehension issues. Tennessee Law Review 67(3 Spring). 701–42.

Dumas, Bethany K. 2000b. U S pattern jury instructions: problems and proposals. Forensic Linguistics: The International Journal of Language and the Law 7(1). 76–98.

English, P. W., and B.D. Sales. 1997. A ceiling or consistency effect for the comprehension of jury instructions. Psychology, Public Policy, and Law 3. 381–401.

Gibbons, John. 2003. Forensic linguistics: an introduction to language in the justice system. Malden, MA and Oxford: Blackwell.

Harris, Sandra. 2001. Fragmented narratives and multiple tellers: witness and defendant accounts in trials. Discourse Studies 3 (1 February). 53–74.

Matoesian, Greg. 1999. Intertextuality, affect, and ideology in legal discourse. Text 19(1). 73–109.

O'Barr, William M. 1982. Linguistic evidence: language, power and strategy in the courtroom. New York: Academic Press.

Stygall, Gail. 1994. Trial language: differential discourse processing and discursive formation. Amsterdam and Philadelphia: John Benjamins.

BETHANY K. DUMAS

Crioulo, Gulf of Guinea

Gulf of Guinea Crioulo (GGC) is a cover term for four creoles lexically based on Portuguese. They are spoken on the three islands of São Tomé (around 100,000 inhabitants), Príncipe, and Annobón (about 5,000 each), of which the former two constitute an independent republic while the latter belongs to Equatorial Guinea.

In the following, the languages will be referred to as Sãotomense, Principense, Annobonese, and Angolar, although they are also known by other names. Sãotomense and Angolar are both spoken on São Tomé, the latter by descendants of runaway plantation slaves in the south of the island.

Previously uninhabited, toward the end of the fifteenth century the islands were settled by Portuguese and their African slaves, the latter of which were brought partly from what is today the coast of Nigeria, and partly from Bantu-speaking areas around the mouth of the Congo. It is in this early period that GGC is assumed to have emerged.

GGC varieties are interesting for a number of reasons, not least because they probably are among the oldest creoles known. They are also typologically unusually distant from their lexifier.

All four creoles coexist in a diglossic relationship with the official languages of the two countries (Spanish on Annobón, Portuguese in the case of the three others) and are used neither in media nor education.

Although influenced by Spanish, GGC does not seem to be seriously threatened on Annobón, where it is the universal vernacular (although all speakers are proficient in Spanish). Angolar is believed to have about 9,000 speakers, but is giving way to both Sãotomense and Portuguese. It is not entirely clear how many speakers use Sãotomense or Príncipense. According to a mid-1990s source, virtually everybody on São Tomé was capable of speaking Portuguese, while only half of the children were competent in Sãotomense, indicating language shift. For reasons yet to be fully understood, this process started earlier and has proceeded further on the smaller Príncipe, and in the 1987 census, less than 16% of the population were reported to speak the language, including virtually no children.

The four GGCs share a number of features on all levels, suggesting a common origin. These include, among others things, a circumverbal sentence negation that can be reconstructed as /na … fa/, the second part of which appears clause-finally. This and other similarities can presumably be traced back to São Tomé, which was the first of the three islands to be settled. Subsequently, however, the languages have drifted apart to a degree where mutual comprehension is difficult, albeit possible.

Like other creoles, GGCs are heavily analytic, and a participial suffix /-du ~ -ɾu/ seems to be the only attested bound morpheme.

The Tense Mode Aspect system is mostly preverbal, and typically includes a past marker *tava* (< P *estava*), an imperfective element *ka* (< P *ficar, cá* or *capaz*?), indicating habitual aspect in isolation, and progressive when preceded by the copula *sa*), and completive particles *za* and *kaba* (< P *já* and *acabar* respectively).

Features that set GGC apart from Portuguese, but make them similar to other creoles in the Atlantic area, include reflexive constructions involving the word for 'body', the use of 3pl as a nominal pluralizer, and use of verb serialization, for all of which African influence may plausibly be invoked.

249

Although being SVO, the GGC languages display several rather exotic orderings from a European perspective, such as N NUM ɔmã dɔs 'two hands').

Principense includes a coarticulated stop /g͡ɓ/, corresponding to /kw/ elsewhere. It is not clear whether nasal+stop sequences should be regarded as cluster or as single phonemes. In all varieties except Principense, Portuguese /v/ turns up as /ɓ/ in some items. Principense and Angolar have a rhotic phoneme, but otherwise Portuguese rhotics appear as laterals or are deleted.

All have depalatalized etymological /ʃ, ʒ/ into /s, z/, but also sport a productive palatalization rule whereby /t, d, s, z/ are realized as [t͡ʃ, d͡ʒ, ʃ, ʒ] before high front vocoids (exceptions: Principense lacks [d͡ʒ], and Angolar has /θ, ð/ in positions where its sisters have /s, z/).

Another segmental difference within the group is that in Annobonese the velar plosive /k/ is realized as [x] before back vowels (including /a/).

The GCCs have been argued to be either tonal or pitch-accent languages. All varieties have a few words that are etymologically consonant-initial, but that have been equipped with a prothetic vowel, possibly deriving from the Portuguese definite article. These items all belong to the core lexicon, and are, interestingly, far more common in Principense (e.g. *ufógo* 'fire') than elsewhere (*fógo*).

With the exception of Angolar (see below), the African lexical contribution is limited to a couple of hundred items at the most. A third of these are derived from languages of present-day Nigeria, most of the remainder being of Bantu origin. The latter group is better represented in Sãotomense than elsewhere. Noteworthy African contributions include 3pl pronouns, and also a 2pl form in Principense.

Not unexpectedly, the Annobonese lexicon nowadays contains a fair proportion of Spanish loans. Due to labor migration to Fernando Poo, where pidgin/creole English is the lingua franca, Annobonese has also assimilated some lexical items from this language.

Lexically, Angolar is the odd man out in the GGC group. While being structurally very similar to Sãotomense, it contains an unusually high proportion of African lexicon, largely derived from Kimbundu.

References

There are five book-length treatments of GGC, listed below. In addition to these, some additional data can be found in scattered articles by Germán de Granda (mostly Annobonese), Marike Post (Annobonese), Tjerk Hagemeijer (Sãotomense) and Luís Ivens Ferraz, and in Valkhoff (1966).

Barrena, Natalio. 1957. Gramática annobonensa. Madrid: Consejo Superior de Investigaciones Científicas.

Ferraz, Luís Ivens. 1979. The creole of São Tomé. Johannesburg: Witwatersrand University Press.

Günther, Wilfried. 1973. Das portugiesische kreolisch der Ihla do Príncipe. Marburg: Marburger Studien zur Afrika- und Asienkunde.

Lorenzino, Gerardo. 1999. The Angolar Creole Portuguese of São Tomé. Munich: Lincom Europa.

Maurer, Philippe. 1995. L'Angolar: Un créole afro-portugais parlé à São Tomé. Hamburg: Helmut Buske Verlag.

Valkhoff, Maurius. 1966. Studies in Portuguese and Creole.Johannesburg: Witwatersrand University Press.

MIKAEL PARKVALL

See also **Crioulo, Upper Guinea**

Crioulo, Upper Guinea

Upper Guinea Crioulo is a continuum of Portuguese-vocabulary creoles spoken mainly on the Cape Verde Islands and in Guinea-Bissau. A variety of Guinea-Bissau Creole also spills over into Casamance, the southernmost province of Senegal. The total number of native speakers is estimated at 400,000 in Cape Verde, 250,000 in Guinea-Bissau, and 40,000 in Senegal. In addition, there are half a million second-language speakers in Guinea-Bissau and some 20,000 in Casamance. There is also an important diaspora of Cape Verdeans in both Europe and the United States.

Despite not having official status in either of the three countries (although since 1998 the 'national language' of Cape Verde), Portuguese Creole nevertheless has a strong position in Cape Verde by virtue of it being the native language of the entire population, and in Guinea-Bissau through being the only nationwide lingua franca.

Judgments on how similar Cape Verde and Guinea-Bissau Crioulo really are differ. While some Cape Verdeans claim to be unable to understand Guinea-Bissau Crioulo, others treat all Upper Guinea

Portuguese Creoles as dialects of one language. In any case, the differences lie more in pronunciation and vocabulary than in structure. Also, Cape Verde Crioulo is in itself rather heterogeneous, and the varieties of the northern Sotavento islands are in most respects intermediate between those of the southerly Barlavento islands and Guinea-Bissau Portuguese Creole. These three main varieties thus constitute a continuum.

Some authors insist that Cape Verde and Guinea-Bissau Portuguese Creole emerged independently, despite the great similarities. Most, however, seem to agree that they are somehow genetically related. Even so, there are two possibilities—did the ancestral Creole travel from the mainland to the islands or vice versa? Both hypotheses have been suggested, without any conclusive evidence being presented for either.

Regardless of the direction of influences during the birth of the Upper Guinea Portuguese Creoles, it is clear that Cape Verde Crioulo has had some impact on the mainland varieties from at least the eighteenth century. First, the Portuguese often used Cape Verdean employees in Guinea-Bissau. Second, famine led to migration toward the mainland in the late 1800s, and third, islanders were prominent in the independence movement.

Anyone accustomed to the better-known Caribbean creoles will notice the—as yet unexplained—absence of many of their characteristic features in the Upper Guinea Portuguese Creoles.

Broadly speaking, the Sotavento varieties (and even more so Guinea Bissau Portuguese Creole) are typologically more distant from Portuguese, and present more features perceived as typical of creole languages. This is manifested e.g. in the use of (mostly) independent words to mark tense. Although usually referred to as a creole, it may be that Barlavento Cape Verde Crioulo should more properly be designated as a semi-creole.

Although there is little evidence of this, many an observer has suggested that the Cape Verde dialects have in general been significantly more creole-like, and that decreolization has brought them closer to Portuguese with time. Today, Many Cape Verdeans are competent in Portuguese, which remains the official language. An obvious difference between the Cape Verde dialects and the mainland is that the latter creole varieties are still in close contact with African languages. At least some change in the direction of Portuguese has taken place during the twentieth century, and is reflected in variation between e.g. more Portuguese-like Guinea-Bissau /adiviɲa/ vs. basilectal /dibiɲa/ 'to guess' (<*adivinhar*) /aʒudar/ vs. /ɟuda/ 'to help' (<*ajudar*).

Lexically, all Upper Guinea Portuguese Creoles are overwhelmingly Portuguese, and only about 150 African items have been attested in each variety, few of which can be considered part of the core vocabulary. In addition to this, there are a number of idioms of seemingly African origin.

Conspicuous phonological features include the replacement of /z/ by /s/ in some varieties, and the realization of Portuguese /ʎ/ as /ɟ/. Furthermore, Portuguese <x> generally turns out as /s/ in Upper Guinea Portuguese Creoles, while instances of orthographic <ch> (/t͡ʃ/ in older Portuguese) become /c/ or /t͡ʃ/. Portuguese syllable structures have, especially in the mainland varieties, drifted toward a Cape Verde structure.

Affixes are scarce, but less so than in most creoles. In Guinea-Bissau, verbs are morphologically marked for causative and passive, in addition to which there are about eight or nine other productive affixes. Comparatives are basically expressed via independent particles across the board, mostly using /mas ADJ (di) ke/ 'more ... than'.

Tense, mood, and aspect are for the most part marked by means of free adverbial markers, including a progressive seemingly derived from *estar* and a past from *jà*. In addition, there is a suffixed *-ba*, probably derived from the Portuguese imperfect *-va*.

As opposed to Portuguese, sentences in the Upper Guinea Portuguese Creoles may not have a subject that is merely implied, although overt subject pronouns frequently attach to the following verb.

Number is normally not marked on nouns, although usually on determiners. Portuguese definite articles have also been lost, with demonstratives taking over some of their functions. The sentence negator is normally not a reflex of Portuguese *não* 'not', but rather /ka/, probably from Portuguese *nunca* 'never'.

In many respects, it is striking, however, how much like Portuguese Creoles are in their syntax. Cape Verde Crioulo replicates the Ibero-Romance distinction between the copulas *ser* and *estar*, relatively faithfully and many varieties also display (variable) adjectival agreement, as in /kaza bnita(a)/ 'beautiful house'. Passives are also constructed very much like in Portuguese. Within the Cape Verde dialects, creole-like structures vary with more Romance-like constructions, as in /kes fiɟu di ʃefi/ vs. /keʃ fiɟ cɛf/ 'the chief's sons'.

Relatively few structural features can reliably be related to African languages. The distinction in Cape Verde Criolo between the two *habeo* verbs /tẽ)/ 'to have, to own' vs. /teni/ 'to have (at hand)' has been suggested to mirror Wolof, one of the African languages with the greatest influence on Portuguese Creoles. Other possible African features include the replacement of etymological /v/ by /b/ in some varieties, and the use of the word 'head' for 'self'.

251

References

Cardoso, Eduardo Augusto. 1989. O crioulo da ilha de São Nicolau de Cabo Verde. Praia: Ministério da Educação.

Couto, Hildo Honório do. 1994. O crioulo português da Guiné-Bissau. Hamburg: Helmut Buske Verlag.

Kihm, Alain. 1994. Kriyol syntax: the Portuguese-based Creole language of Guinea-Bissau. Amsterdam and Philadelphia: Benjamins.

Pinto-Bull, Benjamim. 1989. O Crioulo da Guiné-Bissau: Filosofia e Sabedoria. Lisbon: Ministerio da Educação.

Quint-Abrial, Nicolas. 1998. Dicionário Caboverdiano-Português. Variante de Santiago, CD-ROM edition. Lisbon: Coputação e Linguagem & Priberam Informática.

Rougé, Jean-Louis. 1988. Petit dictionnaire étymologique du Kriol de Guinée-Bissau et Casamance. Bissau: Instituto Nacional de Estudos e Pesquisa.

Scantamburlo, Luigi. 1981. Gramática e dicionário da língua criol da Guiné-Bissau (GCr). Bologna: Editrice Missionaria Italiana.

Silva, Baltasar Lopes da. 1957. O dialecto crioulo de Cabo Verde. Lisbon: Imprensa Nacional.

Wilson, W. A. A. 1982. The Crioulo of Guiné. Johannesburg: Witwatersrand University Press.

MIKAEL PARKVALL

See also **Crioulo, Gulf of Guinea**

D

Dakota and Siouan Languages

The Siouan languages are referred to in more recent literature as the Siouan-Catawban family, as the older term often really meant only the Western Siouan languages, which consists of most of the languages, whereas Catawban and Woccon comprise the Eastern Siouan group. The Siouan-Catawban languages, of which the Dakota languages are one branch, were spoken, at the time of European contact, as far north as Alberta and Saskatchewan, stretching southward through the middle of the United States into Mississippi and Arkansas. Those languages still spoken today are found mainly on reservations in North and South Dakota, Nebraska, Montana, Minnesota, Manitoba, Saskatchewan, and Alberta. There are 20 known languages in all. According to J.W. Powell (1891), in 1890, there were approximately 43,400 officially recognized Siouan tribe members. According to 2000 US Census Bureau statistics, 153,360 people identified themselves as Siouan tribe members, and in Canada, 3,880 people reported membership in a Siouan/Dakota tribe in 2001.

The term 'Siouan' seems to have originated from the shortened colonial French, *Nadouessioux*, an interpretation of an Ojibwa term meaning 'a small rattle snake' or 'the snake-like ones'. In either case, it was a term for a traditional Ojibwa enemy.

The designation 'Sioux' at the time of first contact by Europeans referred to a rather loose confederation of seven tribes—the Sioux nation. The people of these tribes called themselves the Dakota, the Lakota, or the Nakota, which meant 'allies'. These three groups are usually referred to as 'Dakota' and their dialects make up the Dakota languages, which in turn, are part of the Siouan-Catawban languages. The distinction among these three languages was based on what is called the *d/l/n* correspondences; where Dakota has *d*, Lakota has an *l,* and Nakota has an *n*. However, further study has shown that this distinction is inadequate and not necessarily accurate for these languages. Furthermore, Assiniboine and Stoney have been shown to be related to the Dakota languages despite the relationship not having been recognized initially as these two tribes had already split from the Dakotas prior to contact in 1640.

The Siouan-Catawban languages consist of four main branches: a Mississippi Valley group, also called Siouan Proper or Mississippi River; a Missouri Valley group; an Ohio Valley group or Southeastern group; and an eastern group, Catawban. The only language in dispute over its placement is Mandan, which is listed in the Ethnologue as part of the Mississippi group, but others believe that it may stand alone as a subgroup since similarities to either Mississippi group languages or Missouri Valley languages are probably due to borrowing.

Mississippi Valley Languages

The Mississippi Valley languages are divided into four subcategories: the Dakota languages, the Dhegiha languages, the Chiwere languages, and Hochunk. The Dakotan languages were once thought to consist of three or four languages, Lakota (also called Teton), Dakota (also called Santee or Santee-Sisseton), and Nakota (also called Yankton or Yanktonai), but recent work suggests that there are five dialects, including

Assiniboine and Stoney. Total speakers in both the United States and Canada number about 19,000. Dakota (also *Dakhota)* has approximately 4,000 to 5,000 speakers, including children and about 30 monolinguals. There are 250 speakers of Nakota and about 6,000 to 7,000 speakers of Lakota (also *Lakhota)* out of a population of 20,000 according to the 1990 US Census. Assiniboine (also called Hohe) has about 150 to 300 fluent speakers in Canada and Montana out of a population of 3,500 tribal members. Assiniboine shares nearly 90% lexical similarity to Lakota and Stoney, and 90% to 94% similarity to Dakota. Stoney is spoken by 1,000 to 1,500 people out of a population of 3,200 in Canada. Although Stoney shares approximately 90% lexical similarity with Assiniboine, the two languages are not immediately mutually intelligible.

Most of the Dakota tribes have undertaken some measures to preserve their languages or increase the number of speakers. Some Lakota communities are making an effort to teach their children the language since they estimate that by 2013, less than 10% will be able to speak the language if they cannot find a way to stop the language loss. The tribe is working with Indiana University to establish language revitalization projects. The Stoney tribe has also initiated a Language Project, and Dakota and Lakota are taught in schools in North and South Dakota, Montana, Nebraska, Oregon, and Minnesota.

There are four Dhegiha languages: Quapaw, Osage, Kansa, and Omaha-Ponca. Quapaw (also called Arkansas) became extinct sometime in the 1970s or 1980s although there are still nearly 2,000 tribe members. Kansa is also nearly extinct, with just a couple of elderly speakers left. Both languages were last spoken in northern Oklahoma. Omaha-Ponca is considered one language that has barely perceptible differences to nonspeakers. Estimates of numbers of speakers are consistent for Ponca at about 25 fluent speakers, all over 65, but varies widely for Omaha from 20 or so fluent speakers over 60 years old to 1,500 speakers. Most sources agree with an estimate of a couple of dozen speakers. Only a handful of speakers remain for the last language, Osage. The 2000 US Census asserts that over 15,000 people claim membership in the Osage tribe. Speakers and linguists working on the Dhegiha languages have preserved these languages in grammars, dictionaries, and/or texts and a few classes are taught in some of the languages, but it is doubtful that any of these languages will be able to revitalize.

The third branch of the Mississippi Valley group, Chiwere, is sometimes listed as two languages, Iowa and Oto, or as one, Iowa-Oto. The dialects are nearly indistinguishable and are divided on tribal lines rather than linguistic criteria. The Missouria joined the Oto in the late 1700s, so that their dialect, which is believed to have been another language of the Mississippi Valley group, is no longer distinct. The last fluent speaker of Oto died in 1996 and only half a dozen fluent speakers of Iowa remain.

The last language of the Mississippi Valley group was formally known as Winnebago, now Hochunk, 'people of the sacred language'. Winnebago was the name used by the Algonquin, which meant 'foreigners' and adopted by the Europeans. The Winnebago changed their name officially to the Hochunk Nation due, in part, to the association of the name to the RV brand as well as a desire to name themselves. There are about 250 speakers left in a population of about 6,000, but the tribe has undertaken measures to reverse their language loss, including building a Language and Culture Center in Wisconsin with a mission to teach Hochunk to all tribe members, especially the children, from kindergarten through college.

The Missouri Valley Languages

The Missouri Valley group includes two closely related languages, Crow and Hidatsa. Crow is a growing language through efforts of the tribe, including establishing their own two-year college where approximately 75% of students speak Crow as their first language. While classes are not conducted in Crow, the business functions are. In total, there are about 4,500 Crow speakers. Hidatsa, on the other hand, is spoken by just over 100 speakers, all elderly, who live in western North Dakota. There are no efforts to revitalize this language; hence, it will probably become extinct in the next generation.

The Ohio Valley Languages

The Ohio Valley Languages include the extinct languages Tutelo, Ofo, and Biloxi. Ofo and Biloxi were last spoken during the first half of the twentieth century and the last speaker of Tutelo died in the latter half of the century. Tutelo is also sometimes referred to as Saponi, but only a list of Saponi words has survived. They were probably dialects as it has been reported that Saponi and Tutelo speakers could understand one another.

Catawba and Mandan

The remaining languages of the Siouan-Catawban family are the most distant. The Catawban group is usually not listed as a group but a single language, Catawba. However, it is believed that there once were a number of languages that fit into this branch. Adequate materials exist on Catawba, but only a vocabulary list of 143 words in Woccon, a language

that seems to be related by both morphological and lexical properties to Catawba, but these properties are not found in other Siouan groupings. Catawba became extinct when the last speaker died in 1996.

The last language, Mandan, has been thought to stand alone as a branch of the Siouan-Catawban languages while others feel it belongs with the Mississippi Valley group. The resemblances to Crow and Hidatsa are probably due to borrowing, while properties linking it to Mississippi Valley languages were inherited. Mandan is nearly extinct with only six speakers left, all of whom are also Hidatsa speakers.

Siouan-Catawban Features

The Siouan-Catawban languages are generally characterized by Subject–Object–Verb word order. The sound system of the languages is generally typified by the fact that stops ([p, t, k]) have several distinct manners of articulation—unaspirated, aspirated, as well as ejective. The languages have a full inventory of voiceless fricatives, two nasals, *m* and *n*, and the glides *w* and *y*. Most of the languages have three nasal vowels plus five oral vowels, and the vowel length is distinctive.

Morphologically, the Siouan-Catawban languages make extensive and complex use of affixes. Verbs are the most highly inflected category, taking locative, instrumental, reflexive, and pronominal affixes. Case marking is also a feature. Pronominal prefixes are used to indicate first, second, and inclusive persons, while the third person is generally left unmarked or implied.

These pronominal prefixes occur in two sets: one for semantic agents (the entity performing an action) and the other for patients (the entity affected by an action). Particles on the verb may encode a host of other meanings, including plurality and negation. Some languages have such particles for marking word forms that are exclusively used by males or females in commands, assertions, or in giving permission. In these languages, it is also common for nouns to become an integral part of the verb.

References

Boas, Franz. 1991. Introduction to handbook of American Indian languages (1911). Lincoln: University of Nebraska Press.

Chafe, Wallace L. 1976. The Caddoan, Iroquoian, and Siouan languages. The Hague: Mouton.

Mithun, Marianne. 1999. The languages of native North America. Cambridge: Cambridge University Press.

Powell, J.W. 1991. Indian linguistic families of America North of Mexico (1891). Lincoln: University of Nebraska Press.

Rankin, R.L. 1994. Siouan languages. The encyclopedia of languages and linguistics, ed. by R.E. Asher. Oxford: Pergamon Press.

Rankin, Robert, John Boyle, Randolph Graczyk, and John Koontz. 2002. A synchronic and diachronic perspective on 'word' in Siouan. Word: a cross-linguistic typology, ed. by R.M.W. Dixon and Alexandra Y. Aikhenvald. Cambridge: Cambridge University Press.

Rood, D.S. 1979. Siouan. The languages of native America: historical and comparative assessment, ed. by L. Campbell and M. Mithun. Austin and London: University of Texas Press.

Swanton, J.R. 1943. Siouan tribes and the Ohio valley. American Anthropologist 45.

LYNN ANN BURLEY

Deep Structure

The distinction of the *deep* ('underlying') and the 'surface' structure of a sentence reflects the fact that the actual structure of the sentence, in which the verb can be understood as a predicate accompanied by one or more arguments (and adjuncts, adverbials), is expressed more or less regularly by the means of endings, function words (especially prepositions and conjunctions), and word order. Thus, e.g. the English prepositions *in, into,* and *from* carry the functions of Locative, Goal, and Origin, respectively; the surface subject (preceding the verb) expresses the deep subject (Actor) with a verb in the active voice (*Jim saw Mary*) and the deep object with a verb in the passive

(*Jim was seen by Mary*). The distinction of deep and surface structure, already discussed by H.B. Curry (1961), was systematically elaborated in the 1960s, in connection with *constituency*-based grammars (especially with the transformational description of language, initiated by N. Chomsky). Later, this distinction became characteristic of some of the linguistic approaches based on *dependency* syntax. We present here first a brief survey of the position of this level in different stages of theories based on constituent structure (and transformations), and then we outline the present-day role of deep structure in descriptions using dependency syntax, with illustrations taken from

the Praguian approach of Functional Generative Description.

Chomsky's (1965) *Standard Theory* of linguistic description was later *extended* to contain not only a division of the sentence into constituents (such as the verb and its arguments) but also the opposition of initial Phrase Markers (as the representations generated by a context-free phrase structure grammar) and enriched *D-structures*. The transformational component, which took the D-structures as its input, then consisted of the single, general rule 'move alpha', and transferred these structures into S-structures. The word order in an S-structure may differ from the underlying order due to the operation of the rule 'move alpha', i.e. to some of the constituents moving (usually to the left). Traces of the source positions of such movements are indicated in the S-structures by the symbol 't'.

Let us illustrate the framework of the Extended Standard Theory by the following examples:

(1) (a) You told me that Bill saw who?
    (b) Who did you tell me that Bill saw t?
(2) The police know who the FBI discovered that Bill shot t.

While in (1)(a) the *wh*-pronoun occupies its primary position as the direct object of the verb form *saw*, in (1)(b) this pronoun has undergone two movements determined by the cyclic application of 'move alpha'. In both cases, it moves to the beginning of the given clause. Also in (2), the pronoun *who* has moved in a similar way, although here it has not been placed at the initial position of the whole sentence.

Substantial contributions to the discussion on the levels of the system of language can be seen in C. Fillmore's account of '*deep cases*' (such as Agentive, Experiencer, Theme, Instrument), as well as in the controversy between N. Chomsky's 'interpretative semantics' and the 'generative semantics' of G. Lakoff, J. McCawley, and others, or in J. Bresnan's distinction between constituent structure and *functional* structure.

In the recent approach, formulated by N. Chomsky in the 1990s and called *Minimalism*, the derivation of a sentence is characterized as starting by the choice of lexical units, which may then undergo the effects of the 'merge and move' rules (combining words in accordance with restrictions defined in the lexicon, see Abraham et al. 1996). To a certain degree, the position occupied in the preceding versions of Chomskyan linguistic description by the deep structure can be compared to the '*spellout*' point of the derivation; at this point, the representation of the sentence reaches the form from which its phonological form is derived. Another counterpart of deep

```
         prefers
         /  \
       Jim   girls
       /       \
    friend     have
    /         /  \
   my       who  hair
                  /
                dark
```
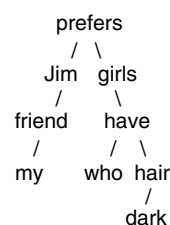
Figure 1. A simplified scheme of a dependency tree for sentence (3).

structure in the minimalist framework is *Logical Form*, which reflects the semantically relevant patterning of the sentence.

The concept of deep structure (partly even with the use of Curry's term 'tectogrammatics') was further elaborated in the context of *Dependency Grammar*, which analyzes the composition of the sentence from word forms, and shows that the sentence is patterned in a highly perspicuous way. The verb can be understood as the center ('head') of the sentence, on which its arguments (such as subject, direct and indirect object, etc.), adjuncts (adverbials of place, time, manner, cause, etc.) depend, since every verb determines a set of such (obligatory and optional) complements. Dependency Grammar emerged in European linguistics in the 1830s (with K. Becker) and penetrated into school teaching in several countries of the Continent. It was systematically elaborated in the 1930s by L. Tesnière from a viewpoint closely related to the Prague School of functional and structural linguistics.

While some of the dependency-based approaches render just a kind of 'surface structure' of the sentence (often found useful in computational linguistics), other such theories work with *deep* structure; cf. especially the frameworks of I. Mel'chuk, J.D. Apresjan, and the Praguian Functional Generative Description (FGD, see Sgall et al. 1986; Hajičová et al. 1998). The sentence is then described as consisting of pairs of lexical *heads* and their *dependents* (modifiers, complements). A typical sentence can be patterned as a dependency tree (a finite graph in which every pair of nodes is connected by a single sequence of edges, and a single node is determined as the root of the tree); every edge connects a head with one of its dependents, cf. Figure 1.

(3) My friend Jim prefers girls who have dark hair. In FGD the tree can be written in a linearized form, with pairs of parentheses used to embrace every dependent:

(3') (((my) friend) Jim) prefers (girls ((who) have ((dark) hair)))

The underlying syntactic functions are rendered by labels of the edges of the tree, or by indices at the

parentheses (at those oriented towards the head), and the values of grammatical categories (tense, aspect, number, etc.) are represented by indices of the labels of nodes. Example (4), whose deep-structure representation (4') displays a small subset of the values, illustrates the handling of the relations of coordination and apposition in FGD:

(4) Mary and Jim, her husband, who live in Boston, have two sons.

(4') $(((Mary\ Jim)_{\text{Conjunction}}\ ((she)_{\text{Appurtenance}}\ husband.\text{Sing})_{\text{Apposition}}\ (_{\text{Descriptive}}\ (who.\text{Plur})_{\text{Actor}}\ live.\text{Present.Indicative}\ (_{\text{Locative}}\ Boston))_{\text{Actor}}\ have.\text{Present.Indicative}\ (_{\text{Objective}}\ (two)_{\text{Restrictive}}\ son.\text{Plur})$

This account reflects the fact that the deep structure requires a descriptive framework of more dimensions than the two that characterize a tree; cf. the groups connected by *and* (conjunction) and by the first comma (apposition) in (4). Such a view differs from that of I. Mel'chuk, who understands coordination as a specific case of dependency.

The syntactic properties of a lexical unit are handled on the basis of its possible dependents, such as Actor (deep subject), Addressee (deep indirect object), Objective (deep direct object), Means, Locative. With nouns, possible dependents are the Restrictive or Nonrestrictive (Descriptive) adjunct, Appurtenance, etc. These kinds of dependencies are expressed by function words, endings, or word order. The syntactic *valency* of a word (its 'valency frame') comprises its obligatory and optional dependents, their properties, restrictions on their combinations, on their movement, etc. Lexical entries thus contain much of the grammatical information.

The distinction between a head and a dependent can be specified on the basis of an operational criterion: the dependent is syntactically omissible, if not in a lexically specified pair of words, then at the level of word classes: e.g. in ((very) slow) *progress,* the syntactic potential of the heads prototypically is identical to that of the whole groups. In *Jim met Sally* nothing can be deleted, but other cases show that the verb is never deletable (without a specific context), whereas object can be absent not only with objectless verbs (*sleep, emerge*) but also with verbs like *read*. Verbs such as *rain* even occur without a lexical subject (*it* as in *it rains* is not a lexical subject, but rather a filler without semantic relevance).

The linearized form of the underlying representations is made possible thanks to their *projectivity* (nonprojective, i.e. discontinuous sentence parts are handled by rules mediating between underlying and morphemic representations). The possibility of linearizing the prototypical structures shows that the

more-dimensional networks can be handled by limited means, similar to those of the propositional calculus. This may be useful for specifying how the core of language can be acquired by children.

A formal definition of deep structure has been formulated by Petkevič (1995) both in the form of a generative procedure and in that of a declarative definition. Other components of a complex description can handle the relationships between underlying syntax, morphemics, and phonology, including the more or less regular expression of grammatical values by morphemes, function words, and specific word order positions, as well as the possibility of a zero expression and other specific phenomena.

The importance of the *interactive* communicative conditioning of language makes it necessary to analyze the sentence with due regard to its position in the context, paying attention to its *topic-focus articulation*. This articulation, which includes the specific syntactic positions of focus-sensitive operators such as negation or *only, even, also* (see Hajičová et al. 1998), reflects the degree to which the structure of natural language has been influenced by the 'given-new' strategy (starting a prototypical utterance by referring to items activated in the hearer's memory).

References

Abraham, Werner, Samuel David Epstein, Höskuldur Thráinsson, and C. Jan-Wouter Zwart (eds.) 1996. Minimal ideas. Syntactic studies in the minimalist framework, Vol. 12. Amsterdam/Philadelphia: John Benjamins.

Chomsky, Noam. 1965. Aspects of the theory of syntax. Cambridge, MA: MIT Press.

Chomsky, Noam. 1975. Reflections on language. New York: Pantheon.

Chomsky, Noam. 1995. The minimalist program. Cambridge, MA: MIT Press.

Curry, Haskel B. 1961. Some logical aspects of language structure. The structure of language and its mathematical aspects, ed. by R. Jakobson, 56–68. Providence, RI.

Hajičová, Eva, Barbara H. Partee, and Petr Sgall. 1998. Topic-focus articulation, tripartite structures, and semantic content. Dordrecht: Kluwer.

Hudson, Richard. 1984. Word grammar. Oxford: Blackwell.

Kahane, Sylvain (ed.) 2000. Les grammaires de dépendence—Dependency grammars. Revue T.A.L., 41, No. 1, Paris.

Mel'chuk, Igor. 1988. Dependency syntax: theory and practice. New York: State University of New York Press.

Petkevič, Vladimír. 1995. A new formal specification of underlying structures. Theoretical Linguistics 21. 7–61.

Sgall, Petr, Eva Hajičová, and Jarmila Panevová. 1986. The meaning of the sentence in its semantic and pragmatic aspects, ed. by Jacob L. Mey. Dordrecht:Reidel and Prague:Academia.

PRAGUE PETR SGALL

See also **Bresnan, Joan; Chomsky, Noam; Lakoff, George**

Definiteness

Primarily a pragmatic, essentially deictic ('pointing at') function, definiteness is expressed cross-linguistically by different devices: phonological, morphological, syntactic, and lexical. The most characteristic such device is the definite article (*the*), i.e. a bound morpheme operating on a noun or noun phrase. When it operates on a nonnominal element, the latter is nominalized—it is turned into a noun. Conversely, all deictics and nominals that function deictically (i.e. all linguistic elements that 'point out' a referent), including proper nouns, are definite intrinsically.

Definiteness is a scalar opposition, i.e. definiteness/indefiniteness are two poles between which there are multiple intermediate points. Nonreferential indefiniteness and denominalization are iconically bound to be marked by zero (indicated below by ø), intermediate degrees are cross-linguistically marked by several devices, e.g. indefinite articles (*a*), a clitic deictic demonstrative (*this-*), etc.

Definiteness is a multidimensional notion that can combine referentiality, specificity, identification, actualization, genericity, individuation, familiarity, and shared knowledge. Some combinations are:

—definite referential, specific, identifying, cf. ***The book*** *I am reading is Tom Sawyer*;
—indefinite referential, specific, nonidentifying, cf. *Tom Sawyer is **a book I** am reading*;
—definite referential, specific, shared knowledge, cf. *I'm looking for **the book** [**I** was reading] #*
—indefinite referential, specific, nonshared knowledge cf. *I'm looking for **a book** [bu:k]… (≈ that was here a minute ago] #*
—indefinite nonreferential, nonspecific, nonshared knowledge cf. *I'm looking for **a book** [buk] # (≈ any book).*

The last two utterances clearly differ by content and context. The first of the two may answer a question of the type *What are you looking for [on the table/in the room/...]?*, or: *Have you lost anything*?, etc. The person answering has a specific book in mind. The second utterance, on the other hand, may represent the first sentence of a client entering a store, who does not necessarily have a specific book in mind. As far as form is concerned, both utterances are likely to differ as well, by means of vowel length, intonation and prosody. In the first one, the accentuated vowel of the indefinite noun is likely to be slightly longer than in the second utterance, where it is non-marked for length. The intonation contour of the first is less clear-cut and the utterance does not end as abruptly as the second, whose intonation contour is the one characteristic of the affirmative sentence, with a clear descent of tone and ending in a clear-cut pause.

In English, if an abstract noun is definite, it is actualized, cf. *ø **Truth** is what we should stand for, but **the truth** is that we don't.* Other nouns whose referents too are seen as nonindividuated, i.e. mass nouns, are incompatible with the indefinite article, cf. ***The / *a sand***. Compatibility is obtained through individuation by numeral classifiers, cf. ***a grain*** *of sand.* When a member of a set is definite but nonreferential, nonspecific, nonindividuated, it is generic, i.e. stands for the whole set and is equivalent to the indefinite nonreferential, nonspecific, nonindividuated plural, cf. ***The bear*** *hibernates* ≈ *ø **Bears** hibernate*. ***A bear*** *hibernates*, in contrast, would be indefinite nonreferential, nonspecific, individuated. Unique elements are definite, e.g. ***the sun***, although they may not be, if they are seen as part of a set, cf. *love under **another sun***. There are languages that devote a special form or syntactic structure to mark the indefinite partially referential, cf. French *Je cherche **du pain** 'I'm looking for **some bread'***.

Negative constructions are hardly compatible with definiteness since most of its dimensions are absent, cf. Fr. *Je veux de **la soupe** 'I want **some soup'*** vs. *Je ne veux pas de ø **soupe** 'I do not want ø **soup'**,* Russian *Ivan kupil **komputer** 'Ivan bought **a computer'*** (accusative) vs. *Ivan ne kupil **komputera** 'Ivan did not buy **a(ny) computer'*** (genitive).This is valid for ergative languages too, cf. Basque *Nik dut baratze bat* 'I have a garden' (absolutive) vs. *nik ez dut baratzerik* 'I do not have a garden' (partitive). If negation is identified contrastively, definiteness is possible, cf. *Je ne veux pas **la soupe,** je veux **la salade** 'I do not want **the soup**, I want **the salad'***.

Nouns that are incorporated into a verb are incompatible with definiteness, cf. *to go hunt **a bear*** vs. *to go ø **bearhunting***, and so are other denominalized nouns, e.g. adverbialized ones, cf. *take ø **fire***.

A particular effect is obtained when definiteness operates on nouns definite by nature, e.g. proper names (of which the definite article is not a permanent constituent)

—referential, specific, cf. *I'm looking for ø **(Mr.) Jones***
—referential, specific, identifying, cf. *I'm looking for **the Mr. Jones** who was here yesterday*

—referential, specific, nonidentifying, cf. *I'm looking for **a Mr. Jones** who is supposed to live here* (when an explicit article is present, prefixed civility classifiers (*Mr...*) or suffixed human classifiers (*...boy/girl*), cf. ***a/the Mr. Jones/Jones boy/guy/Beth girl,*** etc. block the reifying effect of the article).

In English, the definite article allows also to pluralize and actualize proper nouns, e.g. last names: ***The Smiths***. In Córdoba (Argentina) Spanish, in rural French, etc., it is first names that are actualized by the definite article in all functions to convey familiarity. On the other hand, when an indefinite article is used with a proper name, it strips it of its individuality and makes it the representative of a class, cf.: *It would take **a Henry James** to describe that man's psychology*. There are languages in which the definite article operates on proper nouns that are the topic of the utterance. Topicality (old information) and definiteness are narrowly correlated, as are focality (new information) and indefiniteness. In Nêlêmwâ (Melanesian), /-xe/ functions both as a definite article and a topicalizer. Topics tend to be subjectal, agentive, human, deictic, and first actants of transitive verbs; topical nouns with any or some of these properties tend to be definite. Focal (new information) ones tend to be predicative, objectal, patientive, nonhuman, nondeictic, second actants of transitive verbs and indefinite. If definite and/or human, they are discursively marked, and often formally as well, cf. Sp. *Vi **la casa** 'I saw **the house**'* vs. *Vi **a la mujer** 'I saw **the woman**'* Contemporary Hebrew (CH) *[ra?iti ø dira]* vs. *[ra?iti ?et ha-?iʃa]*, Guaraní *[aheʃa oga-ø]* vs. *[aheʃa kuña-me]*. Hence, existential constructions (*There is...*) in which the noun is the focus are cross-linguistically incompatible with definiteness, cf. Spanish **Hay **el libro#** * 'There is **the book#**'* (the asterisk marks ungrammaticality), CH **[yeʃ ha-sefer#]*, Fr. **Il y a **le livre#*** 'id.' One way to override this contraint, viz. to actualize or topicalize an indefinite noun, is to use a deictic demonstrative, cf. *There was **a guy#*** vs. *There was **this guy,** who...* or to focalize the existential, cf. **CH** *[yeʃ ø-sefer#]* vs. *[yeʃ **no ha-sefer#**]*. Conversely, a means to focalize a definite noun is the presentative construction, cf. *Here is **the book,** Fr. Voilà **le livre**,* **CH** *[hine ha-sefer]*, Sp. *He aquí **el libro***. Accordingly, the definiteness gradient correlates with (1) aspect: bounded action ~ definite agent vs. unbounded action ~ indefinite agent; note that genericity blocks the actualizing aspect, cf. *ø **The bear** hibernates /* is hibernating*; (2) dynamicity: active verb ~ definite agent vs. stative verb/adjective/nominal predicate ~ indefinite actant; (3) inherency: operating on a nominal predicate, the indefinite article assigns the subject to a set established by that predicate, cf. German *Die Kirsche ist ø **sauer** 'The cherry is sour'* vs. *Die Kirsche ist **eine***

sauere 'The cherry is of the sour type', Fr. *Il est ø psychologue* 'He understands people' vs. *C'est **un** psychologue* 'He is a psychologist'. The link between (1), (2), and (3) is apparent in Spanish, where *estar* ('be', punctual-dynamic-accidental) is incompatible with the indefinite article, while *ser* ('be', durative-stative-inherent) is compatible with it, cf. respectively **Está / Es una cereza amarga/(un) sicólogo;* (4) noun class, including sex gender. In languages displaying this category, its marks coalesce with those of deixis and often definiteness so that the class prefixes in Bantu; Guaykuru (Amerind); etc., function as definite articles.

Diachronically, a definite article is descended from a deictic demonstrative. Discursively, the definite article is an anaphoric i.e. an intradiscursive deictic device *par excellence*, i.e. it always points to something mentioned, either previously or afterwards, or given/inferable from context (including general truths). This is accomplished either explicitly, cf. *We reached **a river** nearby. **The river** was majestic,* or implicitly, cf. *We reached **a river** nearby. **The other bank** was too distant to be seen.* Deixis is also the first function cast upon the definite article by the child acquiring language. These facts illustrate that definiteness is essentially deictic, and hence of a communicative-pragmatic nature, which is why it is conveyed in all tongues, albeit not necessarily by a specific morpheme. Quintilian's (born AD 35) words: *Noster sermo articulos non desiderat, ideo in alias partes orationis sparguntur* 'Our language does not want articles; hence, thei(r functions) are cast upon other parts of the sentence' apply cross-linguistically; languages not having developed a specific form of a deictic demonstrative to work as a definite article apply to other mechanisms to perform this task. Classical Latin is an example, cross-linguistically current, of definiteness marking in a tongue with no articles: a definite noun is placed in sentence initial position (which often coincides with subject position). There are languages that developed a definite article, then lost it as such either by phonological or by semantic attrition, and then developed a new one. This includes, among others, African languages of various stocks. In Ethiopian-sporadically in Ge'ez and widely in Amharic-a 3rd p. possessive suffix (of deictic origin) is used as a definite article. In Nahuatl, the deictic-nominalizer /in/ functions as definite article when prefixed to the noun; this is corroborated by the fact that when a noun does not bear such a prefix, it is predicative. The suffix /-tl/ marks a vast majority of nouns (except in incorporation, in the plural, when the noun is possessed and in quantifiers, indefinites, and interrogatives); Neo-Aramaic /-a/ behaves similarly. Those are erstwhile deictics that cliticized into definite articles, and then spread to all nouns in all positions and became mere nominalizers.

The numeral 'one' often develops a clitic form to mark an indefinite noun as referential, and the process

starts by marking it as specific: CH [ʔexad/ʔaxat] 'one', respectively, m. and f., evolved a clitic form [-(ʔe)xad/-(ʔa)xat], cf. **[cipor(ʔa)xat *ʔamr-a li*] *'a (certain) bird told me'* vs.[ha-xasida *hi cipor-nod ø*] *'the stork is a migrating bird'*. At present, an anteposed, concording and often stressed form of */ʔejze/* 'which', followed by the relative particle /ʃe/ and a third person deictic is spreading to focalize not the noun itself but its being indefinite referential, specific-, cf. *[jeʃ ʔejzoʃehi hitkadmut]* 'there is some [undoubted] progress]'. Both recent marks are incompatible with each other as well as with the definite article */ha-/* and with a free deictic, which confirms that (in)definiteness is a scalar opposition.

References

Bello, A. [1847] 1984. Gramática de la Lengua Castellana. Madrid: Edaf.

Bril, I. 1994. Indéfini et degrés de définitude en nêlêmwâ. Faits de Langues 4. 211–20.

Corblin, F. 1987. Indéfini, défini, démonstratif. Genève-Paris: Droz.

Givón, T. 1981. On the development of the numeral 'one' as an indefinite marker. Folia Linguistica Historica—Acta Societatis Linguisticae Europaeae I(2). 295–304.

Goldenberg, G. [1991] 1998. On predicative adjectives and syriac syntax. Studies in Semitic Linguistics, 579–90. Jerusalem: Magnes Press.

Grasserie, R. de la. 1896. De l'article. Mémoires de la Société de Linguistique de Paris 9. 285–322, 381–94.

Greenberg, J. 1978. How does a language acquire gender markers? Universals of human language, 47–82. Stanford: Stanford University Press.

Gregersen, E. 1967. Prefix and pronoun in Bantu. IJAL, Memoir 21. Baltimore: LSA.

Guillaume, G. [1919] 1975. Le problème de l'article et sa solution dans la langue française. Paris: Librairie A.-G. Nizet and Québec: Presses de l'Université Laval.

Harris, M. 1979. The marking of definiteness: a diachronic perspective. Papers from the 4th International Conference on Historical Linguistics, ed. by E. Closs Traugott, 75–86. Amsterdam: John Benjamins.

Karmiloff-Smith, A. 1979. A functional approach to child language: a study of determiners and reference. Cambridge: Cambridge University Press.

Kirtchuk(-Halevi), P. I. 1993. Deixis, anaphore, accords, classification: morphogenèse et fonctionnement. Lille: ANRT.

Kirtchuk(-Halevi), P. I. 2000. Deixi and Noun Classification in Pilagá and Beyond. Between Grammar and Lexicon, E. Contini Morava and Y. Tobin (eds.), 31–55. Amsterdam-Philadelphia: John Benjamins.

Kolde, G. 1996. Nominaldetermination. Tübingen: Niemeyer.

Krámsky, J. 1972. The article and the concept of definiteness in language. The Hague–Paris: Mouton.

Launey, M. 1996. Retour au /-tl/ aztèque. Amerindia 21. 77–92.

Leiss, E. 2000. Artikel und Aspekt - die grammatischen Muster von Definitheit. Berlin–New York: Walter de Gruyter.

Lyons, C. 1999. Definiteness. Cambridge: Cambridge University Press.

Rosén, H. 1991. The definite article in the making, nominal constituent order and related phenomena. Linguistic studies on Latin; selected papers from the 6th International Colloquium on Latin Linguistics, ed. by J. Herman, 129–50. Amsterdam–Philadelphia: John Benjamins.

Wackernagel, J. 1928. Vorlesungen über Syntax. Basel: Birkhäuser.

PABLO I. KIRTCHUK-HALEVI

See also **Deixis; Determiner; Reference**

Deixis

Deixis concerns the ways in which languages encode or grammaticalize information regarding the extralinguistic context in which utterances occur, and the ways in which the interpretation of utterances depends on the analysis of that context. For instance, the sentence 'Peter likes me' cannot be properly interpreted unless we know certain details about the context in which it was uttered, namely, the identity of the speaker.

The study of deixis goes back to the work of the ancient Greek grammarians (the term 'deixis' is the Greek word for 'pointing' or 'indicating'). Although deixis has proved itself as a central linguistic concept, during the twentieth century there has been a considerable philosophical interest in deictic phenomena. In fact, the topic was reintroduced into modern linguistics by the German philosopher Karl Bühler in his book *Sprachtheorie* (1934; *Theory of language*). Three main categories of deixis have been traditionally distinguished: person, time, and place. It is generally accepted that deixis is an egocentric system in that the speaker chooses his own perspective when integrating personal, temporal, and spatial information into the message: the central person is the speaker, the central time is the time at which the speaker encodes the message, and the central place is the speaker's location at coding time.

Person deixis involves the identification of the participants in the communicative situation as well as other individuals referred to in the utterance. Among the linguistic expressions that fulfill this identifying function, personal pronouns and inflectional affixes

attached to nouns and verbs are the most usual ones. The basic grammatical distinctions in person deixis are the categories of first, second, and third person: the first person refers to the speaker himself, the second person refers to the addressee, whereas the third person refers to a person or persons who are part of the conversational group but who are neither the speaker nor the addressee. In addition to this contrast, person deictic items typically include information about number and gender. The singular/plural distinction is the primary number opposition in all languages. As regards gender, person deictic items may convey information about the sex (semantic gender) or arbitrary class (grammatical gender) of their referents.

Place deixis deals with the specification of the spatial location of the entities referred to in the utterance. Although locations are most commonly established by reference to that of the speaker, some languages also take the addressee's position as the central place. The basic spatial contrast in most languages, illustrated by the English pair *here* vs. *there*, is that of 'proximal to the speaker' and 'distal to the speaker'. Many languages, however, exhibit a three-fold opposition; Spanish, for example, distinguishes three locative adverbs: *aquí* ('near the speaker'), *allí* ('far from the speaker'), and *ahí* ('a little way from the speaker'). Spatial relations may be expressed linguistically by a wide range of items. The most common perhaps are locative adverbs (*here/there*), demonstrative pronouns (*I will take **that***), and adjectives (***this** house*).

Time deixis involves the identification of the time of the events described in the utterance by reference to the time of the communicative situation itself. Time deixis is commonly treated as a metaphorical extension of place deixis: time is seen as a unidirectional stream in which events are located before (or behind), after (or ahead), or simultaneous with the time at which the utterance is produced. Probably, the most common temporal deictic device is tense, although some languages, like Chinese, lack this verbal category. Languages also rely on temporal adverbs (*now/then*) and demonstratives. Demonstratives are said to be imported from the local domain and used metaphorically in expressions like *this week*, which implies proximity to the utterance time, or *that evening,* which implies remoteness. Finally, most languages have lexical items that specify temporal relations, like the English words *today* or *yesterday.*

In the early 1970s, Fillmore's works brought a renewed interest in the study of deixis and new types were added to the traditional inventory. Among them, the most widely accepted are discourse and social deixis.

Discourse deixis concerns the use of certain expressions within the discourse to refer, not to an extralinguistic entity, but to some portion of the discourse itself. Since discourse develops in time, discourse deixis makes use of terms borrowed from time deixis. For example, the expression *the last paragraph* would be analogous to *last week.* In a similar way, discourse can be thought of as having a spatial extension since any point within it can be located either behind or ahead of a central reference point. Thus, place deictic expressions like the adverbs *above* and *below* or the demonstratives *this* and *that* are commonly used to refer to a preceding or following portion of discourse.

Social deixis involves the encoding of the social status and rank of the participants in the communicative situation and the social and personal relationship between them or between one of them and persons referred to. One of the most obvious manifestations of social deixis is the choice of specific pronominal forms (known as 'honorifics') to refer to the addressee in many European languages: French and German, for example, make a distinction between the familiar terms *tu* and *du* and the polite terms *vous* and *Sie,* respectively. Social status can also be encoded throughout particles and affixes that indicate, as in Japanese, respect or deference. Titles of address (*Mr. President* or *My Lord* in the courtroom) also fulfill this social deictic function. Many languages also make a division between two or more registers in which different vocabulary and syntactic constructions are used depending on whether the communicative situation is seen as formal or informal.

In short, the essential feature of deictic expressions is that their semantic values depend on the real-world context in which they appear. Thus, deixis constitutes a natural link between semantics and pragmatics. However, deixis is also bound to nonlinguistic aspects such as the speaker's attitude or the interaction of grammar and culture and thus it offers itself as a common meeting ground for investigations undertaken from different perspectives such as those of philosophy, cognitive psychology, psycholinguistics, sociolinguistics, or anthropology.

References

Anderson, Stephen R., and Edward L. Keenan. 1985. Deixis. Language typology and linguistic description, Vol. III, ed. by Timothy Shopen. Cambridge: Cambridge University Press.

Bühler, Karl. 1934/1982. Sprachtheorie. Jena/Stuttgart: Gustav Fischer Verlag; as Theory of language. The representational function of language, translated by Donald Fraser Goodwin. Amsterdam and Philadelphia: John Benjamins, 1990.

Fillmore, Charles J. 1971. Santa Cruz lectures on deixis, reproduced by the Indiana University Linguistics Club. Bloomington; Indiana.

Fillmore, Charles J. 1997. Lectures on deixis. Stanford, CA: CSLI Publications.

Jarvella, Robert J., and Wolfgang Klein (eds.) 1982. Speech, place, and action: studies in deixis and related topics. London and New York: John Wiley and Sons.

Lakoff, Robin. 1974. Remarks on this and that. Chicago Linguistic Society 10. 345–56.

Levinson, Stephen C. 1987. Pragmatics. Cambridge: Cambridge University Press.

Lyons, John. 1977. Semantics, 2 vols. Cambridge: Cambridge University Press.

Perkins, Revere D. 1992. Deixis, grammar, and culture. Amsterdam and Philadelphia: John Benjamins.

Rauh, Gisa (ed.) 1983. Essays on deixis. Tübingen: Narr.

LUISA GONZÁLEZ-ROMERO

See also **Bühler, Karl; Fillmore, Charles**

Determiner

A *determiner* is a member of a syntactic category that combines with a noun (e.g. *house, book*) to express information about the reference, quantity, proximity, or gender of the noun. It is a function word that adds grammatical information to the basic meaning of a noun. It may appear in different languages as an independent word (English *a, the*), an affix to a noun (Farsi *kitob-ro,* 'book-the'), a combination of both independent word and affix (Swedish *den vita bilen,* 'the white car-the'), or a whole phrase (as *the dog's* in *the dog's owner*). There are several subtypes of determiners, each with its own particular function.

Basic determiners, or articles (*a, an, the*), give information about the definiteness of a noun. Their use depends on the discourse context. In general, a speaker will use the indefinite articles *a* or *an* to introduce a noun into a conversation or to name one that does not have any specific reference. The definite article *the* introduces a noun that has a unique reference and is familiar or readily identifiable from context. The articles thus play an important role in discourse in signaling the appearance of new referents (with indefinite articles) and maintaining their status as current topics of conversation (with the definite article). Not all languages make use of an overt indefinite article, however. Often there may be only one basic definite article, with bare nouns interpreted as indefinite. In English, the absence of an overt article signals either a mass noun (*rice, butter*) or a generic use of a countable noun, i.e. a noun used to identify a class of objects of a certain type, as *students* in *Students cannot afford to live here.*

The use of the indefinite article *a* in English is complicated by the fact that it is ambiguous with respect to specificity. The sentence *I am looking for a woman with red hair* has two interpretations. The article *a* could signal that the speaker is searching for a type of woman identifiable only by the characteristic of hair color. Or, it could signal that he is in fact searching for a particular woman, whom he already knows. The first interpretation is a nonspecific reading of the article *a*, whereas the second interpretation is a specific reading.

Both interpretations are indefinite. Specificity is thus a different concept from definiteness, although both are encoded in the same determiners in English. In other languages, specificity may be more clearly encoded in determiners that are different from definite and indefinite articles. Or, unlike in English, the concept of specificity, rather than definiteness, may be the key factor in how articles are interpreted.

Demonstratives (this, that, these, those) are determiners whose main function is deictic, or 'pointing out'. They indicate the proximity or closeness of a noun to the speaker. A *proximal demonstrative* identifies a noun as close to the speaker *(this chair)*, whereas a *distal demonstrative* identifies it as distant from the speaker *(that chair)*. The distance can be metaphorical or temporal rather than physical, as in *that situation* or *this week*. Although systems of two degrees of proximity are the most common, some languages may express more. Maori, for instance, identifies three degrees of proximity: close to the speaker, distant from the speaker, and distant from both speaker and hearer.

A secondary use of the proximal demonstratives *this* and *these* is to introduce an indefinite but specific noun that will figure prominently in the conversation that follows—for example, *There was this car blocking the intersection,* or *This guy called for you.* In such constructions, the demonstrative is a stylistic alternative to the indefinite article *a*. As these examples suggest, this is a colloquial use of the demonstrative, and it typically appears in openings to anecdotes and jokes. It is an exceptional indefinite use of the demonstrative, which is otherwise inherently definite when used deictically.

Quantifiers are determiners that express the quantity of a noun, for example, *some, any, most,* and the negative quantifier *no*. Quantifiers interact in complex ways with articles and demonstratives. Some are compatible only with definite interpretations *(all, both, every, most)*, whereas others are inherently indefinite *(some)*. *All* and *both* are the only determiners that can combine with other definite determiners in English, as in *all the students*. None of the quantifiers

may combine with indefinite determiners: *all some students*, *each a student. All* and *both* are sometimes referred to as *predeterminers* and might be better analyzed as a separate category of word.

Possessive determiners include possessive pronouns such as *my* and *your* and entire noun phrases marked with the possessive inflection *-'s,* as in *every man's rights.* Possessive determiners imply definiteness and give specific information about the relationship of one noun with another. Usually, this is a matter of ownership, but it could also be a matter of origin or inclusion. In some languages, possessive determiners can encode a distinction of 'alienable' vs. 'inalienable' possession. In inalienable possession, a noun (typically a body part or kinship term) cannot be separated from the possessor. In alienable possession, a noun can be separated from the possessor, and the relationship of possession can be terminated. This is usually the case with inanimate objects. This very specific information can be encoded in the form of the possessive determiner.

Wh- determiners can take the place of any of the other determiner types if the identity, quantity, or ownership of a noun is unknown and must be questioned: *Which chair? What book? Whose rights? Wh-* determiners thus express unspecific reference in English.

In older grammars, the determiner is not always recognized as a separate syntactic category, or the term may be used only to refer to the basic indefinite and definite articles. Instead, the subtypes of determiners listed above are often treated as completely separate word classes. It is fairly recent that these have been given a unified analysis as members of the more general category of *determiner.* This change is motivated by the fact that they all precede a noun and may substitute for one another but in effect never co-occur in the same phrase. Thus, *the these papers* is ungrammatical. A unified analysis is also supported by the similarities in their meanings and the way in which some imply the definiteness usually encoded by *the.* Given this view, even personal pronouns such as *you* and *we* may be classified as determiners, because they may combine with nouns and signal definite and deictic meanings in phrases such as *you hypocrites* and *we taxpayers.*

The determiner's status as a function word was re-evaluated during the mid-1980s, when syntactic theory began to reinterpret all types of function words as the most important (i.e. 'head') words in a phrase. Before this time, function words were viewed as being dependent on lexical categories such as nouns and verbs. Steven Abney argued in 1987 that the determiner was the head of its own determiner phrase rather than simply a word appearing in a noun phrase headed by a noun. By this analysis, which is known as the DP (Determiner Phrase) Hypothesis and is now widely accepted, a phrase such as *the house* is a DP structure, with the determiner *the* acting as the head of the construction. Most determiners are like *the* in that they must combine with a following nominal category, such as *house.* However, personal pronouns may be analyzed as independent DPs that do not necessarily need to combine with any other category.

The DP analysis is supported by facts about the form, word order, and meaning of determiners in many languages. In particular, the DP analysis is a fruitful means of representing similarities between nominal phrases and full sentences. Noam Chomsky pointed out in the 1970s that a nominal construction such as *the Romans' destruction of Carthage* is in many ways analogous to a corresponding sentence *The Romans destroyed Carthage.* The information carried by the determiner is central to the grammatical interpretation of the noun, in the same way that the information carried by tense and inflectional elements is central to the interpretation of a sentence. Because tense and inflection are assumed to be the head categories of the sentence, it follows that the determiner should be analyzed as the head of the nominal construction. The possessive determiner (*the Romans'*) is the structural parallel of the subject of a full sentence.

In some languages, a determiner may agree syntactically with the noun it accompanies. That is, it may show inflection to indicate that it shares grammatical features such as number (singular or plural) or gender with the noun. Agreement between determiners and nouns is similar to the agreement of the subject and verb in a sentence. Extensions of the DP analysis have therefore suggested that the structure of nominal phrases is very similar to that of sentences.

References

Abney, Steven. 1987. The English noun phrase in its sentential aspect. Cambridge, MA: MIT.

Baltin, Mark, and Chris Collins (eds.) 2001. The handbook of contemporary syntactic theory. Oxford: Blackwell.

Bernstein, Judy B. 2001. The DP hypothesis: identifying clausal properties in the nominal domain. In Baltin and Collins.

Hawkins, John A. 1978. Definiteness and indefiniteness: a study in reference and grammaticality prediction. London: Croom Helm.

Longobardi, Giuseppe. 2001. The structure of DPs: some principles, parameters, and problems. In Baltin and Collins.

Lyons, Christopher. 1999. Definiteness. Cambridge: Cambridge University Press.

Ritter, Elizabeth. 1991. Two functional categories in noun phrases: evidence from modern Hebrew. Perspectives on phrase structure, ed. by Susan Rothstein. San Diego, CA: Academic Press.

Siloni, Tal. 1997. Noun phrases and nominalizations: the syntax of DP. Dordrecht: Kluwer.

Van der Auwera, Johan (ed.) 1980. The semantics of determiners. London: Croom Helm.

SHEILA DOOLEY COLLBERG

See also **Definiteness**

Developmental Stages

Identifying the stages of language development cannot be done independent of the notions of 'comprehension' vs. 'production' and 'competence' vs. 'performance', because, when we claim that a child has acquired 'the past tense' or 'how to tell a structured story', we need to know whether we mean that the child shows evidence of comprehending, say, the difference between past and present tense, or whether the child accurately codes the distinction between them in his or her speech or writing. Beyond this, it is possible that data from neither comprehension nor production accurately reflects the child's mental representations (competence). Performance, in other words, always risks being an inadequate reflection of competence.

Despite these problems, most child language researchers are confident in crediting a child with having acquired a particular aspect of language when they demonstrate it in their production in around 90% of the required contexts. Hence, if a child is using the regular past tense '-ed' ending in nine out of every ten cases where it is required, then we can say the child has acquired it. Given that comprehension almost always precedes productive capacity, this conservative approach is generally justified. There are, though, situations in which children (some more than others) pick up chunks of language, sometimes quite large ones, without really understanding what they mean or how they are constructed. In these cases, production precedes comprehension, and crediting acquisition needs to wait for the child's internal analysis of the chunks.

Notions of general, across-the-board stages of language development are difficult to maintain, even when variations in individual children's rate of development are taken into account. Although writers often refer to 'the one-word stage', 'the multiword stage', etc. these provide little substantive information, actually only serving to trigger the association that, in the first case, we are probably talking about (normally developing full-term) children in their second year of life, and in the second, slightly older children, but still preschoolers. The problem with identifying across the board stages is, firstly, that the range of normal language development is large and shades imperceptibly into language disorder. Secondly, the stages are not discrete, one-word utterances co-occurring with two-word utterances for some time. Thirdly, while some aspects of language development are relatively independent of others (e.g. the acquisition of adjectives, or of consonant clusters), others reflect highly dependent relationships. For example, a child will not be able to advance in developing negation in English without the development of the copular verb and the auxiliary system, including insertion of 'do', since the negative particle (n't) suffixes to these tensed verbs in most instances. (Thus, an early expression of negation such as 'Not fit' becomes 'It doesn't fit.' or 'It can't fit.') The issue of what is related to what is extremely important in discussions of the modularity of language, both in terms of the relationship of language development with nonlinguistic development, and in terms of the internal connectedness of developments within language. This is an area where linguistic theory and first language acquisition research are vitally connected.

Language in Infancy

Although infants under a year old do not yet produce language, they are building mental representations for language and are beginning to comprehend the language(s) around them. In addition, the ability to communicate desires and beliefs by vocalization (including intonation) and both hand, face, and body gesture combined with appropriate eye gaze, develops rapidly in the prelinguistic phase. For those who see pragmatic development as part of language development (some do, some do not), the emergence of gesture and nonverbal vocalization, along with the ability to get and direct the attention of another person and engage in reciprocal turn-taking, are all important milestones of pragmatic competence that are in place prelinguistically.

Much of the language used to a preverbal child is comprehensible from accompanying nonverbal clues (e.g. 'Get me the ball' accompanied by a pointing gesture), and in most cases it is the context, rather than the words spoken, which allows the child to understand. In their own productions, children begin with cries and coos, which give way to 'babbling' at around six months. Babbling is characterized by increasingly stable vocalizations that, over a six-month period, approximate the word structure of the language(s) they are exposed to. Toward the end of the babbling stage, 'proto-words' emerge. These are relatively stable vocalizations that seem to have a recognizable meaning. By the time first words emerge, around the end of the first year, infants show significantly more advanced responses to language than their own productions would indicate.

Deciding when the prelinguistic phase has given way to the linguistic depends on cultural norms, children's control over their vocal apparatus, and normal individual variation. Whether a proto-word is recognized as a word depends on how like a recognizable adult word it is, and the extent to which the adults around the child are willing to interpret them as such. In some cultures, it is the emergence of a particular word (e.g. the word for 'breast') that is seen as marking the emergence of language. Finally, some children seem to experiment freely with early words, while others hang back until they are capable of clearer, more recognizable speech.

Language in the Preschool Years

Many observers have noted that lexical development starts slowly (about one new word per week), but that somewhere around the 50-word mark, many children (but not all) experience a sudden increase in vocabulary acquisition. They then add words at a rate that translates into an average of nine new words per day for the next ten years or more. In production, children generally begin by acquiring names for people and objects in their environment as well as useful interpersonal expressions such as 'hi!' and 'bye-bye', with some children having a greater proportion of such expressions in their early vocabularies than others. Pronunciations are often simplified through the use of strategies such as regularizing the consonant–vowel syllable structure ('wawa' for 'water'), deleting final consonants ('du' for 'duck'), simplifying consonant clusters ('tain' for 'train'), and avoiding difficult (and typologically rare) sounds such as /r/ and /θ/. Words also start out restricted in their reference (e.g. 'teddy' may start out being applied to only the child's own teddy, or only to teddies when they are in the child's home), only gradually coming to have the usage of the adult language, and not infrequently passing through a stage of being overextended in their use (e.g. 'teddy' applies temporarily to all soft toys). Words overextended in production in this way are not usually overextended in comprehension. In comprehension, during the second year of life, children begin to be able to understand a wide range of words for people, games and routines, familiar objects, animals, body parts, and action words. They pick up some of these words on only one exposure.

In languages such as English where word order is the main indicator of who does what to whom, early development is marked by the absence in the production of inflectional and derivational bound morphemes and of function words such as prepositions, auxiliaries, and articles. However, in more richly inflected languages, such as Spanish and Finnish, there is evidence at even the single-word stage that bound morphemes such as verbal inflections are beginning to emerge, and as the single-word stage gives way to the multiword stage, children acquiring such languages advance faster in morphology than those learning languages like English (even though full mastery of these systems may take the learner of Finnish well into adolescence).

Researchers have noted that there is a relatively predictable order of acquisition of functional morphemes in English, from the progressive '-ing', prepositions 'in' and 'on', and plural 's' through to the contractible copula and auxiliary, as in 'He's happy' and 'He's walking', respectively. This order, first explored by Roger Brown (1973), is often used to estimate a child's stage of development. The mean length of utterance (MLU), a measure of productive capacity proposed by Brown, is also used. An MLU of between 1.50 and 2.0 represents Stage 1 of development and is usually attained during the second year. MLU then rises through four successive stages to Stage 5 when MLU reaches 4.0 or above, normally at around three-and-a-half years of age. However, children may vary individually in terms of both the pace and overall rate of development.

As MLU rises, children's utterances not only include more content words but also come to include both function words such as 'he', 'not', and 'a', and bound morphemes such as '-ing', '-ed', and '-s'. In adding bound morphemes in their productions, children also try to rationalize the system by overgeneralizing endings to produce things such as 'breaked', 'goed', and 'deers'. Gradually, however (and it is a slow process), irregular forms such as 'broke' and 'went', which tend to make a brief appearance early in development and then disappear as the overgeneralizations take over, reappear and shut out the earlier forms. As complexity increases, children's capacity for expressing basic sentence types (affirmative declaratives, negative declaratives, interrogatives, etc.) becomes more sophisticated in ways that are difficult to describe independent of a particular theory of syntax. Whether acquisition is seen as the accumulation of constructions or the genetically powered interplay of abstract features and the lexicon will affect how the increasing complexity of a child's utterances is viewed. Gradually, though, the auxiliary system emerges, adverbs are put in the right places, subjects of sentences are expressed, etc. Then, at around three years of age, complex sentences begin to emerge with the successive appearance of complements of verbs such as 'think' and 'know' ('I think you're funny', 'I know how many legs a dinosaur has'), infinitive complements ('I want to go to school'), relative clauses ('That's the teddy that Mummy bought me'), conditionals ('I can if I want to'), and conjunctions ('I have to go 'cos it's time', 'I went to Grandma's and we had tea'.).

Semantic development is intimately bound up with cognitive/conceptual development and it is not clear whether language acquisition introduces new semantic possibilities, or whether it provides coding opportunities for concepts already developed through observation of the world, including the language of others in context. Researchers have suggested that certain overarching semantic relationships appear early as soon as words are combined. These include 'nonexistence', argued to be characterized by the expressions 'no more' and 'allgone'; 'recurrence' ('more noise' and 'nother raisin'); and 'attribution' ('microphone hot' and 'animal book'). Later, more mature semantic notions, such as existence ('There dog') and negation ('No fish'), develop followed by quantificational notions coded by words such as 'some', 'lots of', 'many', 'few', and 'several'.

The preschool period is when conversations can be reliably started and maintained linguistically, when children begin to adapt their messages to different audiences, when miscommunications can be repaired, and when stories begin to have proper beginnings, middles, and ends. Words such as 'here' and 'there','this' and 'that', together with 'I/me', 'you', and 'us', whose references shift depending on who is talking, also begin to be used more accurately.

Language in the School Years

Although much of language development is complete by age five, some syntactic constructions do not appear for many children until they enter school. The passive construction in English ('The book was written by the teacher') is one such. However, it might be noted that in languages such as Sesotho, where the passive is used extensively in speech, it appears much earlier. Irregular past tense forms such as 'fell' and 'brought' may not win out over 'falled' and 'bringed' or 'brang' until at least the age of eight or nine, if ever. Complex sentences with relative clauses and other types of complement begin development in the preschool years, but are not consolidated until middle childhood. The extent to which children learn advanced literary constructions is driven by individual interests and educational expectations. The capacity to read and write and to engage in abstract thought drives both late syntactic development and unlimited vocabulary development.

Around the age of six or seven, the capacity to reflect on language itself, to be metalinguistically aware, emerges. This capacity culminates in the ability, among other things, to define abstract nouns, explain the meaning of idioms and proverbs, and solve complex verbal analogy problems. It is also reflected in verbal humor and, coupled with developments in the ability to 'read' the minds of others, the capacity to exploit complex social uses of language, for example, in narratives.

References

Bloom, Lois. 1993. The transition from infancy to language: acquiring the power of expression. Cambridge, UK: Cambridge University Press.

Brown, Roger. 1973. A first language. Cambridge, MA: Harvard University Press.

de Boysson-Bardies, Benedicte. 1999. How language comes to children. Cambridge, MA: MIT Press.

Fletcher, Paul, and Brian MacWhinney (eds.) 1997. The handbook of child language. Oxford: Blackwell.

Foster-Cohen, Susan. 1999. An introduction to child language development. Harlow, Essex: Longman/Pearson Education.

Hoff-Ginsberg, Erika. 2000. Language development, 2nd edition. Pacific Grove, CA: Brooks/Cole Publishing Company.

Karmiloff, Kyra, and Annette Karmiloff-Smith. Pathways to language: from fetus to adolescent. Cambridge, MA: Harvard University Press.

Nippold, Marilyn. 1998. Later language development: the school-age and adolescent years, 2nd edition. Austin, TX: Pro-ed.

Owens, Robert. 1996. Language development: an introduction, 4th edition. Boston: Allyn & Bacon.

Pinker, Steven. 1994. The language instinct: how the mind creates language. New York: William Morrow and Company Inc.

Pinker, Steven. 1999. Words and rules: the ingredients of language. New York: Basic Books.

Ritchie, William, and Tej Bhatia (eds.) 1999. Handbook of child language acquisition. New York: Academic Press.

Slobin, Dan (ed.) 1985–1997. The cross-linguistic study of language acquisition, Vols. 1–5. Mahwah, NJ: Lawrence Erlbaum Associates.

SUSAN H. FOSTER-COHEN

See also **Acquisition; Acquisition Theories**

Dialectology

Dialectology investigates regionally and socially conditioned linguistic variation, called *regional dialectology* and *social dialectology* (or sociolinguistics). In the history of the discipline of dialectology, the following motivations can be distinguished: normative interests, which prevailed in the sixteenth and seventeenth

centuries (correct standard language vs. incorrect dialect); antiquarian interests, starting in the second half of the seventeenth century, when dialect lexicography began; documentary interests in the nineteenth and twentieth centuries (written records of spoken language, including the collection of fairy tales and phonographic records); linguistic interests, also in the nineteenth and twentieth centuries (language history, local dialect monographs, dialect geography, semantic field theory, and structural and generative dialectology); interests in cultural geography in the twentieth century (extralinguistic interpretations of sound and word geography); and psycholinguistic and sociolinguistic interests, also dating from the twentieth century. Although the psycholinguistic approach tries to discern from the dialect the view of life underlying it, and vice versa—this line of research has not had much of an effect—the sociolinguistic approach deals with broad matters such as language and society, dialect and education, and dialect as a language barrier. Finally, an interest in nonlinguists' views of areal linguistics arose in the last decades of the twentieth century; it is called *perceptual dialectology*.

In the beginning of the nineteenth century, linguists often referred to as *Neogrammarians* tried to prove that the sound system of the languages develops and changes according to rigid and highly regular 'sound laws', the term *law* being understood as in the natural sciences. The perceived regularity of sound changes, they thought, could best be observed in dialects, but as it turned out later, this was not the case.

The comparative method of the nineteenth century had produced two theories with regard to linguistic relationships. One was the naturalistic concept of the family tree theory, advocated by August Schleicher, and—in contrast to this theory—the wave theory of Johannes Schmidt, a student of Schleicher's who maintained that adjacent regions resembled each other most and that differences increased with distance. Research in dialect geography or geolinguistics later revealed that this is largely the picture presented by dialects. In modern terminology, these situations are called *geographical dialect continua*. At their farthest geographical points, dialects may no longer be mutually intelligible, but they will be linked by a chain of mutual intelligibility.

An important aim of dialectology is to delineate dialect areas. Traditionally, this has been done—and still is—by drawing lines on linguistic maps, enclosing one linguistic phenomenon each time and isolating it from its surroundings. These lines are called *isoglosses*. Because isoglosses often do not overlap, Hans Kurath coined the term *heterogloss*, which reflects the conceptual content more accurately. However, more and more modern linguistic atlases no longer show isoglosses or heteroglosses, but rather symbols that reveal transition zones much more clearly and are thus more suited to linguistic reality. In delimiting dialect areas, a qualitative approach must be distinguished from a quantitative one. In the first case, a few features, or often only a single feature, are chosen to differentiate dialects from each other. The more important the chosen feature, the more reliable the dialectal structure obtained. Phonemic features carry more weight than phonetic or lexical ones. Phonemes are the smallest units differentiating the meaning of a language. They are structural linguistic features and thus very stable. In contrast, phonetic features are nonstructural; they show a great deal of variation. The least stable element of a language is its lexicon. A lexical feature is therefore least reliable in differentiating dialects. In a limited quantitative approach, a dozen or more features are selected from any linguistic level—whether it is from the sound, word formation, sentence structure, or vocabulary systems—their borders are calculated, and thicker or thinner lines are drawn according to the number of features they distinguish. Then, the results of the various linguistic levels are compared with one another to see whether they are in substantial agreement or not. In recent years, the use of the computer has made it possible to quantify enormous amounts of data and to present the results on linguistic maps. This branch of linguistic geography, which makes use of highly sophisticated statistical methods, is called *dialectometry* or *dynamic dialectology*. Whereas the procedure involving isoglosses/heteroglosses is a sample of a sample, the dialectometric approach is more objective and exact because it is capable of taking into account all the available linguistic data collected in a certain area. Dialectometry sets off the linear approach of traditional dialectology with an areal one. Its methodological procedure is based on the question of identity or nonidentity of two linguistic forms.

Dialectology has also contributed to linguistic theory. In the field of structural dialectology advocated first by Uriel Weinreich in 1954, mention must be made of the principle of maximum differentiation between two phonemes and the description and comparison of dialectal systems, to which members of the Prague Linguistic Circle have also greatly contributed. By formulating diafeature rules, as attempted by, e.g. Lawrence M. Davis in 1973, taking into account both regional and social variation in language, it has been shown that dialectology added an important dimension to generative phonology, too (see, e.g. Rudolph C. Troike 1971).

Results in dialect geography are normally presented in linguistic atlases. Apart from regional and national linguistic atlases, there are atlases of language groups, such as the Panslavic linguistic atlas. The most comprehensive atlas so far is the *Atlas Linguarum*

Europae, the European linguistic atlas, which covers all six language families and their dialects spoken in the whole of Europe. Also, *Sprachbund* phenomena can be treated in a similar way. A Sprachbund is formed by a group of related or nonrelated languages that show systematic similarities in grammar and lexicon, which cannot be accounted for by direct relationship with a common underlying language. An (not undisputed) example of a Sprachbund comprises the Balkan languages. The production of linguistic atlases today involves the most recent information technology, including speaking linguistic atlases.

Schmidt's wave theory was also applied to social dialects. As early as in 1933, Leonard Bloomfield noted: 'The higher the social position of the nonstandard speaker, the more nearly does he approach the standard language' (*Language*, p. 49). Such situations are called *social dialect continua.*

References

Bloomfield, Leonard. 1933. Language. New York: Holt.
Carver, Craig M. 1987. American regional dialects: a word geography. Ann Arbor: The University of Michigan Press.
Chambers, J.K., and Peter Trudgill. 1998. Dialectology, 2nd edition. Cambridge: Cambridge University Press.
Davis, Lawrence M. 1973. The diafeature: an approach to structural dialectology. Journal of English Linguistics 7. 1–20.
Goossens, Jan. 1969. Strukturelle Sprachgeographie (Structural linguistic geography). Heidelberg: Winter.
Preston, Dennis R. (ed.) 1999. Handbook of perceptual dialectology, Vol. 1 (Vol. 2, 2001). Amsterdam: Benjamins.
Troike, Rudolph C. 1971. Overall pattern and generative phonology. Readings in American dialectology, ed. by Harold B. Allen and Gary N. Underwood, 324–42. New York: Appleton Century Crofts.
Weinreich, Uriel. 1954. Is a structural dialectology possible?. Word 10. 268–80.

WOLFGANG VIERECK

See also **Sociolinguistics; Sprachbund**

Diglossia

In his seminal article in *Word* (1959), Charles Ferguson defined diglossia as:

> 'a relatively stable language situation in which, in addition to the primary dialects of the language (which may include a standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) superposed variety, the vehicle of a large and respected body of written literature, either of an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any section of the community for ordinary conversation.'

Using the examples of Greek, Arabic, Haitian Creole, and Swiss German, Ferguson pointed out several characteristics that are common across all diglossic situations. First of all, in terms of function, there is a strict division of labor between the two language varieties: while the superposed variety or the H(igh) variety is used mostly in prestigious domains such as education and government, the vernacular or the L(ow) variety is restricted to informal domains (e.g. family, neighborhood). Second, although the two varieties are genetically related to each other, the H variety is structurally more complex than the L variety. Third, the H variety is more highly valued than the L variety: although there is a sizeable body of literature written in the H variety, the L variety is rarely used in the written form except in

dialect poetry and advertising. Fourth, given its superior status as the language used in prestigious domains, the H variety also tends to be more standardized than the L variety—grammars and dictionaries are written for the H variety, but not usually for the L variety. Fifth, while the L variety is the language of the home, the H variety is not spoken natively by anyone in the community and has to be learned through schooling. Finally, although the L variety may gradually replace the H variety due to factors such as more widespread literacy and broader communication among different regional and social groups, a diglossic situation is relatively stable—it can persist for centuries or even millennia.

A diglossic situation is likely to come into existence under the following conditions: (1) when literacy in the community is restricted to a small elite; (2) when there is a sizeable body of literature in a language closely related to the vernacular of the community, and this literature embodies some of the fundamental values of the community; and (3) when a considerable period of time elapses following the establishment of these two conditions. Nevertheless, although these social contexts favor the emergence of diglossia, they do not necessarily lead to its development. In fact, it is possible that diglossia may develop from other origins.

Diglossic situations are different from other commonly found language situations in several respects.

In contrast to a diglossic situation, a bilingual situation does not maintain a clear functional compartmentalization of the two varieties. In many Arabic-speaking countries (one of Ferguson's canonical examples), for instance, although colloquial Arabic serves as the basic medium of interaction, Standard Arabic is still the preferred variety for formal purposes. However, in a bilingual community such as Flemish- and French-speaking Belgium, both varieties are used to perform similar functions in formal and informal domains. Furthermore, while a diglossic situation involves two genetically related language varieties (at least according to Ferguson's (1959) definition), two or more unrelated languages often coexist in a bilingual or multilingual community (e.g. English, Chinese, Malay, and Tamil in Singapore).

A diglossic situation is also different from a standard-with-dialects situation. In a community with standard-with-dialects variation, there are those who use the standard language (which is usually the more highly valued variety) in everyday conversation and speak it as their mother tongue. In a diglossic community such as German-speaking Switzerland (another one of Ferguson's examples), the H variety (*Hochdeutsch*) is only learned through formal schooling and is not used as the medium of everyday interaction. On the other hand, in Italy (a standard-with-dialects situation), a considerable number of people speak standard Italian natively and use it in formal as well as informal settings.

Ferguson's article has spurred interest in and stimulated research on this particular type of linguistic phenomenon. In particular, there have been numerous efforts over the years to rework Ferguson's definition of diglossia. In particular, while maintaining the strict functional compartmentalization of the two varieties, Joshua Fishman broadened the definition of diglossia to include genetically unrelated varieties. Thus, according to this broad definition, Spanish- and Guaraní-speaking Paraguay would be classified as a diglossic community, in that the two genetically unrelated varieties function like H and L varieties in diglossic situations. However, some have criticized that this definition dilutes the original meanings of diglossia. Although the Spanish-Guaraní situation in Paraguay resembles, say, the diglossic situation in Arabic-speaking Syria, the two differ from each other in terms of the social history of the code matrix and the social processes that led to their emergence. While the former came into being as a result of colonial contact—in other words, the confluence of two independent sociolinguistic traditions, the latter was derived from the internal functional differentiation within a single sociolinguistic tradition. Furthermore, when language shift occurs in a bilingual community such as Paraguay, it is usually the more prestigious variety or the H variety that replaces the L variety. On the other hand, in the terminal stages of diglossia (in Ferguson's sense), the L variety often displaces the H variety. Synchronically speaking, the two language situations may appear the same. Diachronically speaking, however, they are rather different.

In recent years, scholars have called for a shift in focus in research on diglossia. Some suggest more cross-community studies that examine the origin and the development of various diglossic situations. Another kind of productive research, as Ferguson pointed out in a more recent article, would be studying a given diglossic situation during a period of rapid social and linguistic change.

References

Britto, Francis. 1986. Diglossia: a study of the theory with application to Tamil. Washington, DC: Georgetown University Press.

Ferguson, Charles. 1959. Diglossia. Word 15. 325–40.

Ferguson, Charles. 1991. Diglossia revised. Southwest Journal of Linguistics 10. 214–34.

Fishman, Joshua. 1967. Bilingualism with and without diglossia; diglossia with and without bilingualism. Journal of Social Issues. 23. 29–38.

Hudson, Alan. 1991. Toward the systematic study of diglossia. Southwest Journal of Linguistics 10. 1–22.

Timm, Lenora. 1981. Diglossia old and new: a critique. Anthropological Linguistics 23. 356–67.

ANDREW WONG

See also **Ferguson, Charles Albert**

Discourse Analysis

The term discourse analysis was introduced by Zellig Harris in 1952. In an effort to analyze connected speech and writing, Harris took language description beyond the sentence, examining what language elements might occur next to each other, or in the same linguistic environment.

Originating in the field of sociology in the early 1960s, conversation analysis explores norms and patterns in spoken interactions. Conversation analysis pays particular attention to everyday interactions such as casual conversations, 'chat', and 'ordinary narratives'. Conversation analysts have, however, also examined other interactions such as doctor–patient consultations, legal hearings, news interviews, psychiatric interviews, and interactions in courtrooms and classrooms. One of the aims of the conversation analyst is to avoid prior assumptions or speculations about analytical categories in the analysis of their data. Rather, they look for phenomena that regularly occur and then make that the point of further investigation. Researchers are particularly interested in the sequential structure and coherence of conversations.

The areas of research known as 'contrastive rhetoric' and 'contrastive discourse analysis' compare written and spoken genres in different languages and cultures. Kaplan's work in the mid-1960s is especially important here. Kaplan analyzed the organization of English as a second-language students' essays and found different patterns of organization for students from different cultural backgrounds. While many studies in this are focused on academic writing, studies have also been carried out that examine genres such as economic reports, business letters, and the use of language in everyday interactions from the point of view of discourse structuring, politeness strategies, and reader/writer/speaker orientations.

The Birmingham model of discourse analysis (see Sinclair and Coulthard 1975) proposed a 'rank structure' view of discourse, where discourse is made up of a number of lower level parts, much as a sentence is made up of ranked elements such as the clause, phrase, and word. Researchers recorded a number of British primary school classrooms based on the categories of lesson, and proposed a rank structure for the lesson: transaction, exchange, move, and act. They then drew up rules based on the data to show how the acts combined together to form moves, and the moves formed various kinds of exchange. The Birmingham approach has since been applied to many different interactional settings, such as medical examinations, TV quiz shows, and everyday conversation. The notion of moves has also been used in the area of genre analysis where researchers such as Swales and Dudley-Evans have examined the organizational structure of academic genres such as research articles, theses, and dissertations.

The Australian 'genre school', which originated in Sydney in the 1980s, utilizes the notion of language as a system of choices and views on the social functions of language are important here. Research in this area has examined spoken and written texts such as service encounters, academic essays, casual conversations,

and 'description', 'procedure', and 'exposition'-type texts. In particular, researchers use typical organizational structures and the language that is typically used in the realization of particular genres, to determine the relationship between grammatical and lexical items in texts. More recent research considers how relations of power and ideology shape discourse, and the effects that discourse has upon social identities, relations, knowledge, and beliefs. Critical contrastive rhetoric examines cultural differences in language and communication, viewing these differences as dynamic and situated in relations of power and ideologies.

A number of aspects of language use considered under the heading of discourse analysis are also often discussed under the more general heading of pragmatics. Pragmatics is especially interested in the relationship between language and context. It includes the study of how the interpretation of language depends on knowledge of the world, how speakers use and understand utterances, and how language use is influenced by relationships between speakers and hearers. In a relatively short time, the field of Discourse Analysis has examined issues of structure and coherence, while also placing the discussion within a broader social context. Research provides insights into the organization and interpretation of spoken and written discourse. However, what each of these schools of thought reveals is, in part, a result of the perspective the researchers have taken, and the questions they have asked.

References

Chimombo, Moira, and Henry Roseberry. 1998. The power of discourse: an introduction to discourse analysis. Mahwah, NJ and London: Lawrence Erlbaum.

Connor, Ulla. 1996. Contrastive rhetoric: cross-cultural aspects of second language writing. Cambridge: Cambridge University Press.

Coulthard, Malcolm. 1985. An introduction to discourse analysis, new edition. London and New York: Longman.

Coulthard, Malcolm (ed.) 1992. Advances in spoken discourse analysis. London and New York: Routledge.

Fairclough, Norman. 1995. Critical discourse analysis. London and New York: Longman.

Georgakopoulou, Alexandra, and Doinysius Goutsos. 1997. Discourse analysis: an introduction. Edinburgh: Edinburgh University Press.

Halliday, M.A.K., and Raquiya Hasan. 1976. Cohesion in English. London and New York: Longman.

Halliday, M.A.K., and Raquiya Hasan. 1989. Language, context and text: aspects of language in a social-semiotic perspective. Oxford and New York: Oxford University Press.

Jaworski, Adam, and Nikolas Coupland (eds.) 1999. The discourse reader. London and New York: Routledge.

Kubota, R., and Leaner, A. 2004. Toward critical contrastive rhetoric. Journal of second language writing 13(1). 7–27.

Swales, John. 1990. Genre analysis. English in academic and research settings. Cambridge and New York: Cambridge University Press.

ten Have, Paul. 1998. Doing conversation analysis: a practical guide. Thousand Oaks, CA and London: Sage.

Wodak, Ruth. 1996. Disorders of discourse. London and New York: Longman.

BRIAN PALTRIDGE

Discourse Strategies

The concept of discourse strategy is based on a view of communication as a purposeful social activity. The success of this activity will depend on the application of a series of tactics involving linguistic resources (e.g. sounds, words, grammatical structures, topics, functions) as well as nonlinguistic resources (e.g. gestures, dress, proximity). According to this view, discourse strategies are all those linguistic moves that competent language users make, from several possible choices, in order to achieve their aim in what they consider to be the most efficient, effective, and appropriate way. The linguistic choices that language users make in adopting a particular communicative move may involve aspects related to form, i.e. how to say something, as well as aspects related to content, i.e. what to say. Formal choices include pronunciation, grammar, wording, and textual organization. Content choices basically involve the introduction of particular topics and functions.

Taking into account the cognitive process involved, discourse strategies may be approached from two points of view: (1) as an individual enterprise or (2) as a result of a social, interactional process involving the active participation of more than one individual. According to the first approach, a discourse strategy is an individual's implementation of a mental plan in response to an internal signal that a communicative problem needs to be solved. The second approach considers that discourse strategies can be studied as a result of the joint effort (represented by specific verbal moves) of two or more discourse participants to reach an agreement on their communicative and social intentions.

Besides being problem-based and intentional, discourse strategies are often defined as conscious. This is not a characteristic on which everybody agrees. Thus, it could well be that although at some point a speaker may have consciously decided to make a particular verbal move (for instance, responding to a greeting like *How do you do?* with the same question), this expression forms part of such a frequent communicative routine that it has become an automatic, and fairly obligatory, move. From this point of view, and given the difficulty in having access to the workings of speakers' minds, it might be better to avoid the issue of consciousness and to consider a discourse strategy as a systematic way of using language, which is generally recognized by the members of the speech community as aiming at a particular communicative goal.

One way of classifying discourse strategies is to take into account the communicative goal of the speaker. Thus, it is possible to conceive that human communication is the result of the speakers' mutual agreement to abide by two types of constraints related to (1) their social self and (2) the mode of communication being used. In this sense, speakers' deployment of a particular discourse strategy should be interpreted as an attempt to achieve their goal and, at the same time, attend to one or more communicative constraints. Thus, an utterance like *Can I have a cup of coffee?* can be interpreted as a means for the speaker to obtain a drink while attending to the social constraint of avoiding the impression that the conversation partner is being imposed on to carry out the action. Likewise, the use of a signal like *yeah* or *uh-huh* can be understood as a discourse strategy aimed at coping with the constraint of reciprocity that characterizes face-to-face conversation. The constraints proposed below should be understood not as precise maxims of communicative behavior but rather as scales along which individuals locate their discourse strategies depending on their cultural values and the situation in which they find themselves.

The constraints related to the social self that have been studied in greater detail are (1) presentation of self, (2) size of imposition, (3) social distance, and (4) power. In order to attend to each of these constraints, language users can adopt a series of discourse strategies that are conventionally accepted as effective moves. Let us look at some examples. To 'present' ourselves, we tend to avoid assertiveness by preceding our statements with parenthetical verbs like *I think...* or *I guess...* . In order to mitigate the 'imposition' of a

271

command like *Give me a cigarette*, we may express it by means of the question *Can I have a cigarette?* To attend to the 'social distance' or 'solidarity' constraint, we can address someone by their first name instead of their surname. Finally, it is possible to interpret a request for permission and its subsequent granting or refusal as a discourse strategy to acknowledge the 'power relationship' between the discourse participants. Depending on the context or their personality, the speakers may decide on the way they want to respond to the constraints and, consequently, on the 'discourse strategies' they will adopt. For instance, given a particular situation, it is not at all impossible to think of a speaker who at some point decides to show himself or herself as assertive, imposing, distant, and equal in power to the other speaker.

As for the constraints imposed by the mode of communication, in the case of face-to-face conversation, we can mention (1) topical coherence, (2) turn-taking, and (3) linear structure. To attend to the first constraint, speakers make use of discourse strategies like *by the way*, *like I said,* or *anyway* to introduce, reintroduce, or change a topic, respectively. The turn-taking constraint requires the speakers' deployment of strategies for taking the turn (e.g. *yeah but ...*, *let me just ...*), holding the turn (e.g. fillers like *um* or *well*, repetitions), and yielding the turn (e.g. questions, appealers like *you know, right?*). Finally, the constraint imposed by the fact that we produce and process language linearly, that is, one word after another, requires speakers to adopt discourse strategies like the use of the passive construction to change the information focus of the utterance or the choice from a wide array of connecting words or phrases to express the meaning relationship between one utterance and the following one. Other systems of communication (e.g. letters, e-mail, telephone, visual signals) may have different constraints and consequently may demand different discourse strategies.

References

Brown, Penelope, and Stephen Levinson. 1987. Politeness: some universals in language usage. Cambridge: Cambridge University Press.

Calsamiglia, Helena, and Amparo Tusón. 1999. Las Cosas del Decir. Manual de Análisis del Discurso. Barcelona: Ariel.

Gumperz, John. 1982. Discourse strategies.Cambridge: Cambridge University Press.

Jaworski, Adam, and Nikolas Coupland (eds.) 1999. The discourse reader. London and New York: Routledge.

Kasper, Gabrielle, and Eric Kellerman. 1997. Communication strategies: psycholinguistic and sociolinguistic perspectives. London and New York: Longman.

Stenström, Anna-Brita. 1995. An introduction to spoken interaction. London and New York: Longman.

Titscher, Stefan, Michael Meyer, Ruth Wodak, and Eva Vetter. 2000. Methods of text and discourse analysis. London: Thousand Oaks and New Delhi: Sage.

Van Dijk, Teun A (ed.) 1997. Discourse as social interaction. London: Thousand Oaks and New Delhi: Sage.

Van Dijk, Teun A (ed.) 1997. Discourse as structure and process. London: Thousand Oaks and New Delhi: Sage.

JOSEP MARIA COTS

See also **Discourse Analysis**

Dravidian

A group of 24 languages spoken primarily in the Indian subcontinent make up the Dravidian family of languages. It is the second largest group of languages spoken in India, Indo-Aryan, a branch of Indo-European, being the largest. The other language groups of the subcontinent are Afro-Asiatic and Sino-Tibetan.

Out of 24 languages of the Dravidian group, four major literary Dravidian languages of the Southern Dravidian group are spoken in the southern Indian provinces of Karnataka, Andhra Pradesh, Tamil Nadu, and Kerala. Kannada is the language of Karnataka (40,000,000 speakers), Telugu, numerically the largest, is spoken in the province of Andhra Pradesh (over 75,000,000 speakers), Tamil, the oldest of all Dravidian languages, is spoken in the province of Tamil Nadu (70,000,000 speakers) and Malayalam is spoken in the province of Kerala (30,000,000 speakers). Irula, Kurumba, Tulu, Badaga, Toda, Kota, Koraga, and Kodagu are minor languages of the southern branch. The other minor languages are spoken in the northern and central parts of the Indian subcontinent.

Northern Dravidian languages, Brahui, Kurukh, and Malto are spoken in the province of Baluchistan in Pakistan, Bengal, and Orissa in India, Nepal, and Bangladesh.

Central Dravidian languages are represented by Gondi, Konda, Kuvi, Pengo, Manda, Parji, Kolami, Naikri, Gadaba and Naiki, and Kui. These are nonliterary languages spoken in the provinces of Andhra Pradesh, Madhya Pradesh, and Orissa in India.

The Dravidian languages form completely a separate group of languages, although Indo-Aryan languages have influenced the development of these languages primarily in the domain of phonology and lexicon. However, there have been attempts made earlier to link Dravidian languages to Finno-Ugric and Altaic group of languages.

Speakers of Dravidian languages have lived in the Indian subcontinent for more than 3,500 years and they have recorded history of more than 2,000 years. It was once thought that Dravidians were native to the Indian subcontinent. But, recent archeological, anthropological, and ethnographic studies dispute this claim. Based on the types of agricultural and navigational instruments used by these people, it is now believed that Dravidians must have come from West Africa and the Mediterranean region, and settled in the Indus river basin in present-day Pakistan around 4000 BC. The Dravidian civilization that flourished in and around Harappa and Mohenjedaro region on the banks of the Indus River came to be known as the 'Indus valley civilization'. The period of this civilization is broadly placed between 2500 and 1700 BC. Excavations done at this site led to the discovery of over 3,000 seals. These seals have logo-syllabic and pictographic writings on them. Attempts have been made to decipher these writings with very little success so far.

Typological Characteristics

Phonologically, the Dravidian consonant system consists of obstruents (p. t. ṭ. č, k); nasals (ṇ, n, ñ, m); laterals (l, ḷ); the flap (r(D)); semivowels (y, w); glottal (h), and voiced retroflex continuant (ṟ). The presence of retroflex series (ṭ, ṇ, ḷ, ṛ) is characteristic of the Dravidian consonant inventory.

Dravidian has five vowels (i, e, o, u, a) and the vowel length is distinctive. Word stress is predictable. If a word has short vowels, the initial syllable receives the primary stress. However, if a word has a syllable with long vowels (VV), then the syllable with long vowels receives the primary stress. The syllable pattern is basically of Consonant–Vowel (open syllable) type except that sonorants can close the syllable in some languages. If a root word/stem ends with an obstruent, a default vowel (u) is added to the end of the word; e.g. Kannada *kaaḍ.* root word for 'forest', is pronounced as *kaaḍu.*

Dravidian words are easily segmented into its constituent morphemes; e.g. Kannada *maaḍisikonḍanu* '(he) made (him) (do) (it)' is composed of *maaḍ* 'to do, make' + *is* 'causative' + *koḷ* 'reflexive' + *ḍ* 'past tense' + *an* 'third-person singular masculine' + *u* 'default vowel'. Dravidian words mostly have only suffixes and no prefixes.

Nouns in Dravidian inflect for case, number, gender, and person. Verbs in Dravidian are either finite or nonfinite. Finite verbs are inflected for both tense and subject–verb agreement while nonfinite verbs are not. Finite verbs are marked either for past or nonpast tense.

The default word order in a simple sentence is Subject–Object–Verb. Word order is free except that the verb has to be in the final position. In a typical complex sentence, the matrix clause follows rather than precedes its complement; e.g. Kannada *meeriyu awanige suḷḷu heeḷidaḷu endu (complement clause) jaananu tiḷididdanu (matrix clause)* 'John thought that Mary lied to him'. Similarly, relative clauses precede the noun they modify.

Dravidian languages use 'Brahmi' script, which has an alpha-syllabic writing style with diacritics used for vowels occurring in postconsonantal position.

References

Andronov, M.S. 1970. Dravidian languages. Moscow. Nauka Publishing House.
Bhat, D.N.S. 1978. Pronominalization. Poona: Deccan College.
Bhat, D.N.S. 1998. The adjecival category. Amsterdam and Philadelphia: John Benjamins Publishing Company.
Burrow, T., and M.B. Emeneau. 1961. A Dravidian etymological dictionary [DED]. Oxford: Clarendon Press.
Caldwell, R. 1961. A comparative grammar of the Dravidian or South Indian family of languages. Madras: University of Madras.
Emeneau, M.B. 1962b. Dravidian and Indian linguistics. Berkeley: Center for South Asian Studies, UC (Miimeographed).
Emeneau, M.B. 1967b. Dravidian linguistics, ethonology and folktales: collected papers. Annamalainagar: Annamalai University.
Krishnamurthy, Bh. 2003. The Dravidian languages. London: Cambridge.
Masica, Colin Paul. 1976. Defining a linguistic area: South Asia. Chicago: Chicago University Press.
Steever, Sanford B. 1988. The serial verb formation in the Dravidian languages.
Steever, Sanford B. (ed). 1988a. The Dravidian languages. London and New York: Routledge.
Zvelebil, K.V. 1990a. Dravidian linguistics: an introduction. Pondicherry Institute of Linguistics and Culture.
Zvelebil, K.V. 1970b. Comparative Dravidian phonology. The Hague: Mouton.

CHANDRA SHEKAR

Dutch

Dutch is one of the Western Germanic languages, a branch that belongs to the Germanic subgroup of the Indo-European family of languages. It stands about midway between English and German (both geographically and lexically) and is the closest to English of all the major languages of the Germanic family.

Dutch is an official language in the Netherlands and Belgium. It is the mother tongue of more than 21 million people in the Netherlands, in several provinces of Belgium called Flanders (Ghent, Brabant, Antwerpen, and Limburg), in the former Dutch colonies in America (Surinam) and Asia (Indonesia), and in the Netherlands Antilles in the Caribbean Sea.

In Belgium, the language is called Flemish (*Vlaams*), although there is practically no difference compared with Dutch in the Netherlands. Dutch and Flemish are actually the same language, but because of long-lasting cultural and religious reasons, there are two separate terms for one and the same language. The Flemish areas have been subject to French influence, especially from 1830 onward, when the state of Belgium came into being with French as the sole official language. In 1921, Belgium was divided into two linguistic sectors. Flemish became the official language in East Flanders, West Flanders, Antwerp, Limburg, and eastern Brabant, and French became the official language in all the other provinces. This linguistic frontier was adjusted in 1962 and 1963, and the two principal areas were made formally unilingual, but the Brussels region has remained bilingual (the Belgian capital was originally a Dutch-speaking town). However, the requirement for unilingualism applied only to the so-called *buitendienst* 'foreign affairs', i.e. it became compulsory in communications with officials and in official life in general, with a free choice for 'domestic affairs'.

The ancestor of Dutch was the West Low Franconian dialect of Germanic tribes. The history of the Dutch language (whose origin has been set by linguists at approximately 700 CE) may be divided into three main periods—Old, Middle, and Modern Dutch. Old Dutch extends to approximately 1150; there are few direct records of this language, but the only important extant monument of this period is a translation of the Psalter. Middle Dutch extends from 1150 to 1500 and is characterized by a significant influence from Latin and French and by the first Dutch dictionaries. In contrast to Old Dutch, a large number of Middle Dutch texts, both literary and official, have survived. Even from the time before 1300, more than 2,000 texts have been preserved. A large number of these texts are to be found in the *Corpus Gysseling*, which is a standard edition of Middle Dutch texts from the period before 1300.

In the sixteenth and seventeenth centuries, a new linguistic consciousness developed in the Netherlands. During this period, Modern Dutch began to form on the base of the dialect of the Amsterdam region after it had become the capital of an independent nation. In 1584, the first Dutch grammar appeared, written by the rhetorician H.L. Spieghel, the so-called *Twespraack vande Nederduitsche Letterkunst*, 'Dialogue of Dutch Letter Craft', presented in the form of a dialogue (*tweespraak*).

Modern Standard Dutch is the direct heir of the dialects spoken in the provinces of North and South Holland and Utrecht. Starting in the nineteenth century, this standard Dutch increasingly replaced the dialects of other provinces, even in oral use, especially among the more cultivated classes, and it was also adopted in Belgium.

Despite the rather limited geographical distribution of the Dutch language, there are a wide variety of regional dialects whose mutual intelligibility is often low. Traditionally, linguists divided Modern Dutch dialects into six large groups: (1) the central-western dialects, including all those in the provinces of North and South Holland and Utrecht, large parts of Gelderland, and the Zeeland Isles; (2) the northeastern dialects (often called 'Saxon') in Groningen, Drenthe, Overijsel, and the eastern part of Gelderland; (3) the central-southern dialects in the Netherlands province of North Brabant and adjacent parts of Limburg, and in the Belgian provinces of Antwerp, Brabant, and East Flanders; (4) the southwestern dialects in the Belgian province of West Flanders, French-Flanders, Zeeland, and the islands of Goeree and Overflakkee in the province of South Holland; (5) the southeastern dialects (often called 'Eastern Low Franconian') in Belgian and Dutch Limburg and some villages in north Brabant; and (6) the northwestern dialects in North Holland above the IJsselmeer, the non-Frisian Wadden islands, the coastal stretch of Holland province, and the South Holland islands apart from Goeree and Overflakkee.

Dialects that are geographically widely separated from each other sometimes show very significant differences: a speaker of the Groningen dialect will

scarcely be able to communicate with a speaker of West Flemish, unless they both speak standard Dutch. For most Dutch and Flemish speakers, the most evident division is into northern and southern dialects—to the South of the 'great rivers' (the Rhine and the Meuse), people use a 'soft *g*' that is not found in the North. For dialectologists, the contrast between East and West is more important because the western dialects still show a number of Coastal Germanic characteristics.

The Modern Dutch vowel system is the result of a great number of changes, but the Proto-Germanic consonant system has generally survived in Dutch.

The history of the Dutch language witnessed constant simplification in the domain of word structure. The noun system in Middle Dutch displayed three genders: common, neuter, and feminine. Today, Dutch nouns may be either common gender nouns or neuter nouns, with the former taking the definite article *de* and the latter *het*: *de man* 'the man', *de vrouw* 'the woman', *het kind* 'the child', *het huis* 'the house'. The feminine gender has been lost.

In the pronominal system, most pronominal categories make the distinction between two forms according to the nature of the concepts they stand for. Generally, there is a 'neutral' form, used for singular nonhuman referents, and one used for human referents, either plural or singular. Personal pronouns are the only Dutch words that still display a clear opposition between subject and object forms, i.e. Dutch pronouns distinguish between nominative and accusative case. However, there is no difference between 'accusative' (direct object) and 'dative' (indirect object) uses of the object forms.

Dutch word order is generally subject–object–verb; i.e. in basic main clauses, the verb is always in the second position, and it usually follows the subject. The verb appears at the end of the clause in subordinate clauses. It comes at the beginning of the clause in yes/no questions and imperative clauses and in main clauses preceded by a subordinate clause.

The Dutch language has known a long literature tradition, and its lexicon remains quite rich. Moreover, it is quite pure, because Dutch did not borrow many words from other languages. The languages that influenced Dutch over the centuries were French and its northern dialects (especially Picardian), mainly in Belgium, and classical and medieval Latin. Words of Germanic origin form the overwhelming majority of the noncomplex part of the basic vocabulary.

The Dutch spelling system (orthography) is based on the rules devised and published by the Dutch linguists De Vries and Lammert Allard Te Winkel in 1864 and was adopted officially both in Belgium and in the Netherlands in the second half of the nineteenth century. Its authors had used the criterion of 'received

pronunciation'; i.e. generally Dutch orthography closely reflects pronunciation.

Most loanwords come from English, not just as a matter of new technical concepts, but also of loans in varied areas of everyday life. In many cases, the spelling and pronunciation of the English words are adapted to fit Dutch usage. There are also many German loanwords in Dutch. Some of them are already so old and so adapted to Dutch usage that most people do not even realize that the words are of German origin (e.g. *tijdschrift* 'newspaper'). Other German loan-words have managed to retain their original sound and spelling, e.g. *sowieso* 'anyway'. Nowadays, the German influence has weakened because of the strong influence of English. Alongside many French loanwords, there are Dutch equivalents that could not be driven out completely and are now used in parallel. However, the French word is often preferred: *jus d'orange/sinaasappelsap* 'orange juice'.

Dutch had contacts mainly within the Germanic group of Indo-European languages. Contact of Dutch and native languages gave rise to creoles in the Dutch East Indies (as the so-called language of the Sinyos and Nonahs 'gentlemen and ladies'), in the former Danish Antilles (whose colonists were mainly Dutchmen), and in South Africa.

A language that has evolved directly from Dutch is *Afrikaans,* one of the two official languages of the Republic of South Africa, the other being English. It is spoken by more than 4 million people of the various ethnic groups found in the country. Afrikaans is also one of the two official languages of Namibia, a former protectorate of South Africa.

Like Dutch, Afrikaans is also a member of the Western Germanic languages. It developed in the seventeenth century from Dutch brought to South Africa by the first settlers from Holland. Originally, it was a popular dialect composed of Dutch with several borrowings from aboriginal languages of Africa (especially from the Khoisan family and the Niger-Congo family). By 1900, Afrikaans was emerging as an independent language, and in 1925 it was finally recognized as an official language, having been used in the schools and church since 1914 and permitted for use in universities since 1918.

At the level of pronunciation, there is an appreciable difference between Afrikaans and Dutch. Where Dutch uses *oo*, Afrikaans often has *eu,* and Dutch *a* and *aa* were replaced by *e*. Afrikaans is almost the only Germanic language that uses nasal vowels, especially before *s*.

The Afrikaans noun, displaying a pattern similar to the Dutch noun, fully lost the gender distinction.

The Afrikaans verbal system is characterized by the complete loss of person and number. In order to indicate tense, Afrikaans uses auxiliary verbs and adverbs.

During its history, Afrikaans was significantly influenced by the Zulu, Bushmen, and Gottentote languages of Africa, mainly with respect to vocabulary. The drastic simplifications in word structure may also be the result of lengthy contact with African, Portuguese, and Malayan languages as well.

References

Brachin, Pierre. La langue néerlandaise. 1977. Brussels: Didier; as The Dutch language: a survey, translated from the French by Paul Vincent. Cheltenham: Thornes, 1985.

Burridge, Kate. 1992. Syntactic change in Germanic: a study of some aspects of language change in Germanic with particular reference to Middle Dutch. Amsterdam; Philadelphia: John Benjamins.

Donaldson, Bruce. 1997. Dutch: a comprehensive grammar. London, New York: Routledge.

Konig, Ekkehard, and Johan van der Auwera (eds.) 1994. The Germanic languages. London, New York: Routledge.

Koster, J. 1975. Dutch as an SOV language. Linguistic Analysis 1.111–36.

Lippi-Green, Rosina L., and Joseph C. Salmons (eds.) 1996. Germanic linguistics. Amsterdam,Philadelphia: John Benjamins Publishing Company.

Shetter, William Z. 1984. Introduction to Dutch: a practical grammar, 5th edition. Leiden: M. Nijhoff.

LAURA DANILIUC

See also **Belgium; Indo-European 2: Germanic Languages**

Dyslexia

There are two broad categories of dyslexia: acquired and developmental. 'Acquired dyslexia' is a failure of the ability to read that sometimes follows a stroke or other condition that causes injury to the brain. Lesions of the parietal lobe and occipital regions, in particular, sometimes result in severe reading and/or writing difficulties. Often, there is some recovery, particularly if remedial training is given. The more common use of the term 'dyslexia' refers to developmental reading disability. 'Developmental dyslexia' is the failure to learn to read despite receiving instruction that is normally adequate. 'Dyslexia' has sometimes been applied to people who are nearly incapable of reading the simplest text. However, the term is more often used inclusively, of individuals who may read but do so with difficulty in comparison to their peers. A dyslexic, then, is someone whose level of reading ability, as measured by standardized tests, is substantially below the expected level for his or her age and level of schooling. For example, a child whose spoken language is English and who is of adequate intelligence and education but who reads below the 25th percentile on a standardized reading test may be considered dyslexic.

A precise definition of dyslexia depends on whether it is based only on a reader's performance relative to his or her peers (as in the example above), or on a discrepancy between the individual's reading performance and his or her general intellectual functioning. As an example of the latter, a conservative definition of dyslexia has been suggested for individuals whose reading test scores are more than 1.5 standard errors lower than their IQ scores (about a 75% difference) and who have at least the minimal intelligence needed to benefit from instruction in reading (e.g. an IQ of 80). However, discrepancy definitions are problematic. For example, a reader who has a high IQ but reads only moderately well would deviate from expectations of what should have been achieved, given his or her strong intelligence. Such a person would be classified as dyslexic by this formula in spite of being functionally literate. Further, definitions based on discrepancies between reading achievement and IQ have little theoretical or practical significance. It is adequate to define dyslexia only in terms of the reading performance itself relative to reading performance norms for the reader's age (omitting any reference to an IQ discrepancy).

A persistent myth is that dyslexics see letters and words backward. This belief is without foundation. Although reversals can be observed in many beginning readers, these reversals are caused by faulty learning, not by disordered visual perception. Few of these children become dyslexic. Current understanding of dyslexia is that it is not primarily a visual but a language-based disorder, particularly a disorder of sound perception and production. In order to read print in an alphabetic writing system (such as English, Russian, or Arabic—but not Chinese or Japanese), good readers must be able to read words they have never seen before. In order to do this, they must understand the alphabetic principle: that letters correspond to sound units or segments, the fundamental components of speech. To do this requires, in turn, an ability to understand that spoken words are sequences of those sound

segments. With this understanding, a reader can 'sound out' (or 'decode') a printed word, thereby identifying it. Children who are dyslexic nearly always lack an ability to understand consciously the segmental nature of the spoken word. Tests designed to assess a child's ability may ask questions like, 'Can you say *monkey* without the *mmmm*?' or 'Which word has more sounds in it, *car* or *cart*?' The child who cannot respond appropriately is not able to consciously analyze the spoken form. Such a child is at an obvious disadvantage in learning spelling-to-sound correspondences.

Dyslexia is less prominent when the written language's spelling represents the spoken word in a simple, unambiguous manner. For example, the writing systems for Finnish, Russian, and Spanish (to name only a few so-called 'transparent' alphabetic orthographies) have an unambiguous relationship between each sound unit and each letter of the alphabet. Nearly all children who have adequate intelligence learn to decode in these languages. In contrast, so-called 'opaque' writing systems like English and Hebrew have a great deal of ambiguity in letter-to-sound relationships. In English, for example, there are many letters (in particular the vowels), each of which corresponds to several sounds. For example, the consonant letter *g* has three different pronunciations in the words *giant, get,* and *night*. Moreover, there are many sounds that can be represented by different spellings, e.g. compare the identical vowel sound in the words *chief, beef,* and *leaf*.

However, even if a child is able to sound out printed words, she or he will not necessarily become a skilled reader who can read words fluently (i.e. in a rapid, automatic manner) and comprehend well what is read. Even in writing systems that are not opaque or even alphabetic, nonfluent readers exist. Good comprehension in reading depends critically on being able to recognize printed words rapidly. Slow word recognition requires so much of the reader's attention and short-term memory that there is little left over to devote to the cognitive processing demands of comprehension (which includes understanding both grammatical information and meaning). For alphabetic writing systems, slow decoding is seen as the primary bottleneck that blocks good comprehension.

Current evidence from brain imaging research indicates that dyslexics are deficient in certain brain areas believed to be involved in sound processing, visual processing, and the coordination of the two. In spite of congenital individual differences in neurobiology, initial studies suggest that children who are identified early with neurological deficits in brain areas affecting their reading ability can be helped. Most useful is an intensive program of reading instruction that emphasizes the segmental nature of the spoken word and its correspondences to the writing system. This general approach is popularly called 'phonics', although, within that rubric, there are many different methods that differ in detail and effectiveness.

References

Fletcher, Jack M., Robin Morris, G. Reid Lyon, Karla K. Stuebing, Sally E. Shaywitz, Donald P. Shankweiler, Leonard Katz, and Bennet A. Shaywitz. 1997. Subtypes of dyslexia. Foundations of reading acquisition and dyslexia, ed. by Benita A. Blachman, Mahwah, NJ: Lawrence Erlbaum Associates.

Katz, Leonard, and Ram Frost. 1992. The reading process is different for different orthographies: the orthographic depth hypothesis. Orthography, phonology, morphology, and meaning, ed. by Ram Frost and Leonard Katz. Amsterdam: Elsevier North-Holland Press.

Liberman, Isabelle Y., Donald Shankweiler, and Alvin M. Liberman. 1992. The alphabetic principle and learning to read. In Shankweiler and Liberman.

Olson, Richard, Helen Datta, Javier Gayan, and John DeFries. 1999. A behavioral–genetic analysis of reading disabilities and component processes. Converging methods for understanding reading and dyslexia, ed. by Raymond Klein and Patricia McMullen. Cambridge, MA: MIT Press.

Rayner, Keith, and Alexander Pollatsek. 1989. The psychology of reading. Englewood Cliffs, NJ: Prentice-Hall.

Shankweiler, Donald P., and Isabelle Y. Liberman (eds.) 1992. Phonology and reading disability. Ann Arbor: The University of Michigan Press.

Shankweiler, Donald P. 1992. How problems of comprehension are related to difficulties in decoding. In Shankweiler and Liberman.

Shaywitz, Sally A., Bennett A. Shaywitz, Kenneth R. Pugh, Pavel Skudlarski, Rob K. Fulbright, R.Todd Constable, Richard A. Bronen, Jack M. Fletcher, Alvin M. Liberman, Donald P. Shankweiler, Leonard Katz, Cheryl Lacadie, Karen Marchione, and John C. Gore, 1996. The neurobiology of developmental dyslexia as viewed through the lens of functional resonance imaging technology. Neuroimaging, ed. by G. Reid Lyon and Judith M. Rumsey. Baltimore: Brookes Publishing.

LEONARD KATZ

E

Emeneau, Murray Barnson

Growing up in Lunenberg, a linguistically unique small town in Nova Scotia, studying French, German, Latin, and Greek, it is easy to comprehend how a bright Canadian lad was destined to become a linguist. Combining a background in the classics, general linguistics, and anthropology, Murray Barnson Emeneau became one of the greatest linguistic Indologists of all time. A prolific author, no linguist interested in India can afford not to read him, while every general linguist will profit from his lucid prose, no matter what the field of specialization. Among his most noteworthy contributions are his work on the Indian *Sprachbund* or 'linguistic area', bilingualism and structural borrowing, linguistic prehistory, as well as Dravidian comparative linguistics and etymology.

Emeneau's overall approach to linguistics was influenced by the Boasian and Sapirian traditions, which consider it a major branch of cultural anthropology. His anthropological background was largely self-taught through voluminous reading, auditing courses taught by Edward Sapir while still at Yale, and associating with him from 1931 to 1935, and through three years of fieldwork on nonliterary languages in British India. Since academic positions were hard to secure during the years of the Great Depression and were virtually nonexistent for Indologists, Emeneau went off to India on a fellowship from Yale, the American Council of Learned Societies, and the American Philosophical Society. There, he did fieldwork from 1935 to 1938 in the Nilgiris on three little-known Dravidian languages: Toda, Kota, and Kodagu, as well as in central India, where he studied Kolami.

Thoroughly versed in Greek, Latin, Sanskrit, and comparative Indo-European at Dalhousie University, Halifax, Nova Scotia, Oxford University, and Yale, Emeneau brought with him a superb background for his (1931) Ph.D. dissertation, an edition of a Sanskrit folk tale text. His rigorous training under teachers Franklin Edgerton and Edgar H. Sturtevant, together with his experience in linguistic fieldwork, led to his innovative research of the Indian linguistic area — a topic that was totally neglected. Although the notion of the *Sprachbund* was known from the Caucasus, the Balkans, and the northwest coast of the United States (American Indian languages), Emeneau's pioneering investigation of the Indian scene describes a situation in which 'languages belonging to more than one family show traits in common which do not belong to the other members of (at least) one of the families' (Emeneau 1980:127). Franz Boas and Edward Sapir had discussed the problems of the diffusion of linguistic traits across genetic boundaries, using data from various American Indian phyla, and Emeneau used their methodology to investigate the historical relationships of the three major subcontinent language families coming in contact: Indo-Aryan, Dravidian, and Munda (Austro-Asiatic).

Lexical borrowing has long been recognized, and all the Dravidian languages have loan words from Sanskrit and later Indo-Aryan languages. The Munda languages have been influenced by Indo-Aryan vocabulary, and Indo-Aryan languages have borrowed Dravidian words. The relationship among the three families is a direct reflection of the linguistic history and prehistory of India.

Two phonological features used by Emeneau are illustrative of the linguistic diffusion described above. Most Indian languages have a set of retroflex consonants that contrast with dentals. A situation of pre-Indo-Aryan and pre-Dravidian bilingualism allowed pre-Indo-Aryan allophones to diffuse as redistributed retroflex phonemes. Furthermore, Emeneau demonstrated that the Peshawar dialect of (Iranian) Pashto has been Indianized in the occurrence of the retroflexes as well as in borrowing many words from modern Indo-Aryan. Since the retroflexes are Proto-Dravidian and not Proto-Indo-European, this is a clear case of phonological influence, or as Emeneau put it, 'the Indianization of the Indo-European component' (1980:111).

Another example of phonological influence concerns the palatals. In Marathi, the palatals of Old Indo-Aryan affricated into *č* and *ǰ* before front vowels and *ts* and *dz* before back vowels. As a similar development occurs in Indo-Aryan Oriya, in two Dravidian languages, Telugu and Kannada, and in the Munda language Kurky, one can conclude that all of these languages are in contact and form a continuum across central India.

One of the most interesting phenomena used by Emeneau to illustrate areal diffusion is the case of numeral classifiers. The research demonstrates that Indo-Aryan is closer to Dravidian than to Indo-European. Further, the Magadhan languages of modern Indo-Aryan seem to be the origin of all Indian classifier systems.

Two other areal traits deserve mention. First, Dravidian and Indo-Aryan have similar onomatopoeic systems, each with stems that occur in reduplicated and nonreduplicated forms and verbal derivations therefrom. Since it is doubtful that the Indo-Aryan system derives from that of Proto-Indo-European, Dravidian influence is likely, which would account for the Rigvedic occurrences. Second, the North Indo-Aryan verbal system of intransitive, transitive, and causative is reminiscent of the Dravidian causative system.

As for Emeneau's comparative Dravidian work, the genetic relationship among the members of this closely related family (more like Romance than like Indo-European [1980:74]) was of interest to him from the very beginning, and etymologies were easy to discern. Emeneau correctly theorized that the relationship among Latin, English, and Proto-Indo-European was mirrored by that of Kolami, Telugu, and Proto-Dravidian. Emeneau's research on the Kolami lexicon may serve as illustrative: 720 words had good Dravidian etymologies; however, 166 of these were found to be borrowings from Telugu. Desiring to produce a definitive etymological dictionary of this neglected family, Emeneau traveled to England in 1949 and proposed a comparative Dravidian dictionary to England's leading Indologist, the late Thomas Burrow of Oxford University. In many ways, Emeneau's *magum opus*, the coauthored *Dravidian etymological dictionary* (Oxford: Clarendon Press), first published in 1961, 2nd ed. in 1984, capped a long and as yet unfinished distinguished career.

Biography

Murray Barnson Emeneau was born in Lunenburg, Nova Scotia on February 28, 1904. He received his BA (1923 Dalhousie University and 1926 Oxon.), MA (1935 Oxon.), and Ph.D. (1931) from Yale University. He was a Rhodes Scholar in 1923–1926; Governor-General's Gold Medallist 1923; Guggenheim Fellow in 1949–1950, 1956–1957; Faculty Research Lecturer, UC Berkeley in 1955–1956; and Wilbur Lucius Cross Medallist, Yale Graduate School in 1969. He was an instructor of Latin in Yale in 1926–1931; he carried out anthropological linguistic research in India on Dravidian languages in 1935–1938. He was Assistant Professor of Sanskrit and General Linguistics at UC Berkeley in 1940–1943; Associate Professor, 1943–1946; Professor, 1946–1971; and Founder and Chairman, Dept. of Linguistics, UC Berkeley, 1953. He received honorary degrees (L.H.D.) from the University of Chicago in 1968; LL.D. from Dalhousie University in 1970; D.Litt. (Honoris Causa) from University of Hyderabad, India in 1987; and an Honorary Degree, Vidyavachaspati from K.S.D. Sanskrit University, Darbhanga (India) in 1999. He was President, Linguistic Society of America in 1949; Editor, *Journal of the American Oriental Society* (*JAOS*)1941–1951; President, AOS, 1954–1955; Honorary Fellow Royal Asiatic Society in 1969; Fellow of the American Academy of Arts and Sciences in 1970; Corresponding Fellow, British Academy in 1993; Honorary Member Linguistic Society of India 1964 (winner Citation and silver plate, Golden Jubilee of Society, 1978); and Honorary Member, The Philological Society in 2000. He also received a Medal of Merit, AOS, in 2000.

References

Emeneau, Murray B. 1944–1946. Kota texts.
———. 1951. Studies in Vietnamese (Annamese) grammar.
———. 1955. Kolami: a Dravidian language.
———. 1971. Toda songs.
———. 1980. Language and linguistic area.
———. 1984. Toda grammar and texts.
———. Coauthor (with Thomas Burrow). 1961, 1968. A Dravidian etymological dictionary, 2nd edition 1984.
———. 1994. Dravidian studies: selected papers.
———. 1980. Author's postscript. Emeneau, pp. 350–4.

———. 1991. A Nova Scotian becomes a linguistic Indologist. First person singular, Vol. II, ed. by E. F. K. Koerner, 85–101. Amsterdam: John Benjamins.

Krishnamurti, Bhadriraju. 1969. Comparative Dravidian studies. Current trends in linguistics, Vol. V, ed. by Thomas A. Sebeok, 309–33.

Van Nooten, Barend. 1988. Sanskrit Studies of M. B. Emeneau: Selected Papers.

ALAN S. KAYE

See also **Dravidian; India; Sprachbund**

Emotion and Language

There is no established definition of emotion. The term is usually used to refer to feelings, such as joy, jealousy, or disappointment, but it is also used to refer to more diffuse types of moods, such as irritability or stress. Furthermore, it is not clear how many emotions exist. Many disciplines have addressed the issue, and results diverge considerably. Some accounts from psychology list more than 500, but other accounts assume a much smaller number.

Expressing emotions is one of the most fundamental functions of language. This function has been termed 'emotive' (or 'expressive'). Although it may not be the dominant function in many cases, the emotive function affects virtually all utterances.

Emotions may be expressed at all linguistic levels. Semilinguistic noises, such as *wow*, *ouch*, or *tut-tut*, are traditionally classified as interjections. Interjections are considered purely emotive signs. They do not have any cognitive meaning. A conventional feature on the phonetic level is the emphatic prolongation of a vowel, as in [bi:g] for *big*, as opposed to neutral [big]. Intonation and other phenomena, such as word stress, also play an important role. On the word level, the emotive function is expressed, for example, by the so-called diminutive suffixes and familiarity markers, as in *doggie*, *weirdo*, *fats*, and *babykins*. Other word-formation devices include reduplication, e.g. *teeny-weeny*, *okey-dokey*. Emotion can also be expressed by stylistically marked word choices, e.g. *bean* for head. Further lexical means include terms of endearment and swear words. On the sentence level, we find intensification and emphatic repetition (e.g. *This is very very ugly*, *He worked and worked*) and exclamations such as *What a sight!* On the discourse level, speech acts such as insults are important means of communicating emotions. Ritualized communicative patterns are used to react to social standard situations concerning illness, death, birth, etc.; i.e. they are expressions of grief, sadness, commiseration, or happiness.

A distinction can be made between words that express emotions and words that evoke emotions. The latter type can be illustrated by examples such as *mother*, *brother*, *friend*, or *hometown*. Furthermore, utterances may evoke an emotion in the addressee; e.g. a derogatory remark may cause an outburst of fury.

Emotions can also be expressed paralinguistically. Relevant paralinguistic features involve higher pitch, increased tempo, a trembling or a creaky voice, groaning, giggling, and laughing. It is worth noting that different ways of laughing may express different emotions, e.g. an embarrassed laugh, a derisive laugh, or relieved laughter.

Problems in verbal planning and delivery can also signal emotion. Examples include hesitation, repetition, stammering, and unfinished sentences. On the level of turn-taking and discourse organization, phenomena (potentially) indicative of emotion involve interruptions, intentional overlap, simultaneous speech, and—as an extreme form—shouting somebody down. On the level of topic selection and topic treatment, emotions may motivate the choice of taboo topics, or speaking the plain truth.

Emotions can be communicated not only linguistically and paralinguistically but also nonverbally. Nonverbal behavior of this type includes, first and foremost, facial expression. Contrary to popular belief, facial expression conveying a particular feeling is not universal, but culture specific.

Apart from studying verbal manifestations of emotions, linguistics also investigates the way we speak about emotions: when, how, and why humans talk about emotions and how emotions are negotiated in discourse. Emotion words and emotion concepts are the focus of semantic analysis.

Every language has an emotion lexicon, i.e. a set of words that function as labels for emotions. These are used to name and communicate about emotions. The emotion lexicon consists predominantly of nouns, e.g. *fear*, *disgust*, and *pride*, and adjectives, e.g. *sad*, *angry*, and *happy*.

Emotions can be positive, e.g. *joy*, *love*, or negative, e.g. *anxiety*, *shame*. Many emotion words, both nouns

and adjectives, form antonymous pairs, i.e. pairs of opposing emotions. Examples include *joy–sorrow*, *love–hate*, *happy–sad*. Some emotion words have more than one antonym, e.g. *joy–sorrow/grief*, and some emotion adjectives have two opposite partners, one that is related in form, e.g. *happy* vs. *unhappy*, and one that is related only in meaning, e.g. *happy* vs. *sad*. Antonyms related in form are the result of negation, i.e. of adding a negative prefix (e.g. *un-*). In this respect, the adjectival emotion lexicon displays certain asymmetries. For instance, *happy* can be negated (>*unhappy*), but *sad* cannot be negated (>**unsad*). As a rule, only positive emotion adjectives can be negated.

Emotion words form so-called lexical fields. These are subsets of the vocabulary of a language whose members are close or related in meaning. Emotion words of the same field name the same emotion or very similar emotions. For example, the word field ANGER comprises, among other words, *anger, rage, fury, wrath, ire, irritation, frustration, annoyance*, and *indignation*. The most typical representative of this field is *anger*. Less typical members include *annoyance* and *indignation*, whereas *bitterness* and *resentment* are considered only peripheral members. Some members refer to different intensities of the same emotion; e.g. *rage* is more intense than *anger*, and *fury* is even stronger than *rage*. *Annoyance*, on the other hand, is weaker than *anger*.

In the emotion lexicon, some words are considered more basic than others, because some emotions are considered more fundamental than others. This has been established empirically in both linguistics and psychology. The most basic emotions seem to be love, fear, anger, happiness, and sadness. When asked to enumerate emotions, native speakers of English are most likely to mention these five. There is a high degree of similarity concerning these emotions across languages, including not only Western European languages but also typologically distant languages. At the same time, no direct correspondence exists between the emotion vocabularies of different languages. For instance, the same lexical field may have a different number of members in different languages. For one English emotion word, there may be two or more possible translations in another language.

The way we speak about emotions reveals the way we think about them. Cognitive semantics demonstrates that we use metaphors for conceptualizing emotions. Emotion concepts are reflected in proverbs, idioms, and collocations, such as, for instance, *You make my blood boil* or *He's just letting off steam*. Such expressions illustrate that the central metaphor for anger is 'heat of fluid in a container'. Metaphorical expressions of other emotions show that the human body is generally seen as a container for emotions, as in *She was filled with love*, or *Her feelings overflowed*. To a large extent, however, emotion concepts are culture specific.

In summary, linguists examine how we communicate emotions and how we communicate about them. Emotions can be expressed linguistically, and also paralinguistically, at all levels of language. The emotive function is one of the most fundamental functions of human language. The emotion vocabulary is structured in lexical fields, involving basic vs. peripheral words as well as antonymous pairs. There is no one-to-one equivalence of emotion vocabularies across languages. Emotion concepts are reflected in culture-specific metaphorical expressions.

References

Athanasiadou, Angeliki, and Elzbieta Tabakowska (eds.) 1998. Speaking of emotions. Berlin: Mouton de Gruyter.

Fussell, Susan (ed.) 2002. The verbal communication of emotions: interdisciplinary perspectives. Mahwah, NJ and London: Erlbaum.

Kövecses, Zoltán. 1990. Emotion concepts. New York and Berlin: Springer.

Niemeyer, Susanne, and René Dirven (eds.) 1997. The language of emotions. Amsterdam: Benjamins.

Russell, James, Fernández-Dols, José-Miguel, Anthony Manstead, and Jane Wellenkamp (eds.) 1995. Everyday conceptions of emotion. Dordrecht: Kluwer.

Volek, Bronislava. 1987. Emotive signs in language and semantic functioning of derived nouns in Russian. Amsterdam and Philadelphia: Benjamins.

Wierzbicka, Anna. 1999. Emotions across languages and cultures. Cambridge: Cambridge University Press and Paris: Editions de la Maison des Sciences de l'Homme.

KLAUS P. SCHNEIDER

See also **Metaphor**

Empty Categories

The elements of a sentence do not always have to be pronounced. This is true in elliptical structures, in which elements have been left out for stylistic reasons or because they are easily recoverable from the context. After a statement such as *She did something weird*, a possible follow-up question would be *She did*

something weird?, but most speakers would prefer the truncated form *She did? Something weird* remains implied. Some linguists claim that sentences may also contain covert or null elements that are not the result of optional ellipsis, but always remain unpronounced. These are called *empty categories*. The reason for assuming the existence of such empty categories is that they contribute to the meaning of the sentence and that they can also play an important syntactic role. The task of the listener is to identify all these empty categories in a given sentence and to determine their proper function and interpretation. The task of the linguist goes beyond mere identification: a proper formal theory of language has to determine the specific properties of the contexts where these empty categories can appear.

In the early 1980s, Noam Chomsky proposed a classification of empty categories that are important for sentence structures. He identified several different types of empty categories and argued that they can be differentiated according to whether or not they are anaphoric and whether or not they are pronominal. Anaphoric elements refer back to elements that have been mentioned before and require an antecedent, typically in the same sentence. Pronominal elements function in a manner similar to pronouns such as *I*, *she*, or *they*, i.e. they may refer to entities that must be recoverable from the context but do not need to occur within the same sentence.

Consider first an infinitive subordinate clause, indicated by square brackets: *I want [to leave]*. Like all clauses, these must have a subject, i.e. someone who leaves. In this case, it is clear to the listener that the subject of the main clause (*I*) is also the subject of *leave*. However, it is not the case that the speaker could optionally insert *I* into the subject position of the subordinate clause; *I want I to leave* would be ungrammatical, and *I want me to leave* sounds somewhat strange. Thus, the empty subject is not the result of a stylistic type of ellipsis. Rather, Chomsky argued, this is the case of an empty subject that he called 'PRO'. The empty subject of the infinitive has to exist in the sentence structure, because it plays a part in the interpretation of the clause.

The element PRO has both pronominal and anaphoric properties. The fact that PRO is anaphoric means that, for its interpretation, it depends on another nominal phrase. In other words, PRO is 'controlled' by another nominal phrase. For example, in *John wanted [PRO to leave immediately]*, the referential properties of PRO depend on those of the controller, namely *John. John* can thus be considered the antecedent of PRO, and *PRO* is 'controlled' by *John*. There are two types of control: subject control and object control. In the first type, PRO depends for its interpretation on the subject of the main clause, as in *John promises his mother [PRO to study]*. In the second type, PRO has the same referent as the object of the main clause: *Mary forces her son [PRO to eat]*. Whether PRO is controlled by the subject or by the object depends uniquely on the properties of the main verb.

The fact that PRO is also pronominal means that the interpretation of PRO can be the same as a pronoun. In *It is better [PRO to take the car]*, PRO may have the same interpretation as *they* in *It is better [that they take the car]*. In sentences like *It is impossible [PRO to learn the secret code]*, the meaning of PRO is somewhat more vague and may be similar to *somebody* in *It is impossible [that somebody learns the secret code]*; PRO is said to have an arbitrary interpretation in such types of sentences. In its pronominal use, PRO is not controlled by an antecedent.

The second empty category is 'small' *pro*, the subject of an inflected verb. In languages such as Spanish and Italian, the pronoun subject can be left unexpressed, as in the Italian *Ha mangiato* '(He or she) has eaten'. There, the implicit subject is represented by *pro*: *pro ha mangiato*. 'Small' *pro* has to be present in such sentences because, just like large PRO, it represents the subject of the sentence and is thus crucial for its interpretation. Furthermore, it can act as an antecedent for anaphors. For example, the Italian reflexive *si* 'self/each other' in *pro si vedono spesso* '(They) often see each other' is anaphoric and has to rely on another sentence element for its interpretation. The only possible antecedent for the reflexive *si* in this sentence is the null subject of the verb.

The *pro* phenomenon is far from being unitary: sometimes *pro* acts exactly like a pronoun; it has a very specific interpretation, as in the above examples. Sometimes, however, *pro* is used nonreferentially, in contexts where the subject appears in postverbal position: *pro arriva un uomo* '(There) arrives a man'. In these cases, it acts as the nonovert counterpart of the English expletive *there*, i.e. a sentence element that is almost void of meaning. The notion of expletive *pro* has been criticized. The most important objection against this element is that it is postulated on purely theory-internal grounds, namely, the assumption that sentences require a subject *in front of* the verb. However, there is no empirical evidence to justify the existence of expletive *pro*.

Not all languages use *pro*. The languages that allow the pronominal subject to be unexpressed are called 'pro-drop languages' or 'null subject languages'. Since the 1970s, considerable work has been done to determine how *pro* is formally licensed in languages that use it. Many different hypotheses have been advanced,

but the idea present in almost every hypothesis is the 'richness of the verbal morphology'. It has often been said that *pro* appears in those languages that have a rich verbal morphology, i.e. in those languages in which the verb inflects for person, number, and gender. If the form of the verb already provides such clues about the subject, an empty subject is interpretable. Thus, in Spanish *pro hemos comido* 'we have eaten', *pro* is interpretable because the verbal inflection *-emos* clearly indicates that the subject must be first-person plural, i.e. 'we'.

The third type of empty categories depends on the assumption that the construction of certain (or all) sentences involves moving some elements around. For example, sentences like *The man has been bitten by the dog* are often claimed to involve movement of *the man* from the object position of the verb (see Figure 1).

The reasoning is that even though *the man* is in the grammatical subject position, the passive verb requires it to be interpreted as the object. Chomsky proposed that such movement leaves a 'trace', usually indicated by *t* (see Figure 2).

This type of movement trace is anaphoric, because its interpretation depends on *the man*. Because *the man* acts as the antecedent within the same clause, these traces are not pronominal: pro-

nouns do not allow an antecedent within the same clause.

Another type of trace is that left by question words, which are fronted in English (see Figure 3). *What*, here, is clearly not the subject of the sentence: *you* is the subject. Rather, *what* must be interpreted as the object of *done*.

Thus, Chomsky assumed that question formation involves moving the question word to the front of the sentence, again leaving a trace. Because *what* acts as the antecedent of the trace, the trace again cannot be pronominal. However, it cannot be anaphoric either, because it allows its antecedent to occur beyond the immediate sentence boundary, as in Figure 4.

The embedded sentence *that the famous painter will paint t* thus contains a trace without an antecedent within the same clause, something that typical anaphoric elements do not allow.

In short, there are four types of empty categories, PRO, *pro*, traces of regular noun phrase, and traces of question words. These four categories can be formally distinguished as follows: (1) PRO is anaphoric and pronominal; (2) *pro* is pronominal, but not anaphoric; (3) traces of regular noun phrases are anaphoric, but not pronominal; and (4) traces of question words are neither anaphoric nor pronominal.

Twenty years of linguistic research have thus provided a systematic classification of empty categories, and empty categories belong to the standard descriptive tools of linguists working in the Chomskyan tradition. However, one must keep in mind that competing theoretical frameworks do without empty categories or movement. Empty categories are thus theoretical metaphors that may or may not have a correspondence in the actual processing of sentences by human speakers and listeners.

has been bitten **the man** *by the dog*

Figure 1

The man *has been bitten* **t** *by the dog*

Figure 2

What *have you done* **t**?

Figure 3

Whom *do [you think* **t** *[that the famous painter will paint* **t** *]]?*

Figure 4

References

Chomsky, Noam. 1981. Lectures on government and binding. Dordrecht: Foris Publications.

Chomsky, Noam. 1982. Some concepts and consequences of the theory of government and binding. Cambridge, MA: MIT Press.

Haegeman, Liliane. 1991. Introduction to government and binding theory. Oxford and Cambridge, MA: Blackwell; 2nd edition, 1994.

Roberge, Yves. 1990. The syntactic recoverability of null arguments. Kingston: McGill-Queen's University Press.

Ruzicka, Rudolf. 1999. Control in grammar and pragmatics: a cross-linguistic study. Amsterdam: John Benjamins.

KAREN LAHOUSSE

See also **Chomsky, Noam; Generative Grammar**

Empty Morphemes

The concept of morpheme was defined by the first structuralists as the minimum linguistic unit containing signifier and meaning, that is, as a formal element with a content that cannot be divided into smaller segments that, in turn, also possess meaning: *preexistence* contains three meaningful morphemes (*pre-*, *-exist-*, and *-ence*) and cannot be divided into smaller meaningful parts.

The morphemes are recognized by their recurrence, that is to say, for their presence with the same meaning and the same form in different words: in the previous example, *pre-* also appears in *pre-history* or *pre-establish; exist* appears in *exist-ent* or *exist-ential*; and *-ence* in *prefer-ence* or *compet-ence*.

However, some linguists observed that this conception is not always sufficient, because not all the segments obtained by means of formal word analysis meet these characteristics. This can be observed in different situations:

1. The same content can be present in phonetically different segments: the regular English plurals have the forms /s/, /z/ o /iz/ (*hits* /ˈhits/, *boys* /ˈboiz/, *houses* /ˈhausiz/), depending on the circumstances.
2. A segment may not be formally divided in to smaller units, but it may seem to contain more than one meaning: in the lat. *amo*, the segment *-o* cannot be formally divided, but it contains the meanings for 'first person', 'present' and 'indicative'.
3. Some morphemes are linked to the extreme such that the boundary between them disappears, like it is the case of the Spanish *al* (*a + el*) or the French *au* (*à + le*).
4. Some morphemic contents appear in certain words, although they cannot be assigned to any part of the words: the difference between 'present' and 'past' observed in *I beat John every day* vs. *I beat John yesterday*, does not correspond to any formal segment of *beat*.
5. The formal analysis sometimes uncovers recurrent segments that are not necessarily assigned a meaning, such as the case of *-mit -* in the series formed by *remit, demit, commit…*

These problems affect the two faces that the morpheme, as a linguistic sign, possesses: it affects its signifier (problems 1–4) and meaning (problem 5). In order to try to solve them, the linguistic theory has formulated diverse ones. Among them, we are now interested in the following concepts:

On the one hand there is the established distinction between morpheme as the unit of a language, and morph, its phonetic realization. The separation of these two concepts allows to postulate that a biunivocal correspondence among them does not have to exist, in such a way that a morpheme can be represented by several morphs (called allomorphs; problem 1); a morph can comprise more than one morpheme (a phenomenon denominated syncretism; problem 2); the morphs of two morphemes can merge phonically (something usually known as amalgam or *morphs portmanteau*; problem 3); or a morphologic content may not correspond to any morph (it is said to be a zero morph; problem 4).

With respect to problem 5, which affects the meaning of the morpheme, two positions have been adopted that, although they are not always distinguished in the linguistic bibliography, is convenient to differentiate:

In the first place, some authors, adhering to the statement of Leonard Bloomfield that all production of the language is completely composed of morphemes (*principle of total accountability*), would likely affirm that elements like those that appear in the series *re-mit*, *de-mit*, *com-mit*, *trans-mit*; *re-fer*, *de-fer*, *pre-fer*, *in-fer*; *or cran-berry*, *boysen-berry*, *huckle-berry*, in spite of lacking a meaning definable in linguistic terms, they should be considered morphemes, because they appear in a recurrent way in diverse words (the prefixes *de-*, *re-*, *con-*, etc. and the roots *mit* and *fer* are present in many words of the English language) or simply because they appear linked to forms that appear in some other constructions (*cran -*, *boysen-*, and *huckle-* only appear in these three cases, but *berry* appears scattered and it combines with other morphemes in *straw-berry, blue-berry or black-berry*). Consequently, these would be morphemes lacking meaning, or *empty morphemes*.

Such an interpretation comes from Bloomfield himself, who would affirm that if a complex linguistic form contains a residual, it should necessarily be another linguistic form (as in the case of *cran-berry*), and that forms such as *remit* or *conceive* should be interpreted as complex words despite the fact that, in them, the meaning of the morphemes is impossible to

describe. This affirmation would be adopted by later linguists, as the case of Zellig Harris, Eugene Nida, Robert Robins, or Mark Aronoff.

The main consequence of this method of analyzing data is a substantial modification of the concept of morpheme. According to many modern authors, this unit can only be defined as the minimum unit of grammatical analysis, that is, a unit that appears recurrently in the words and that allows the differentiation of some words from others. A paradigmatic example of this idea is the definition by Aronoff: 'A morpheme is a phonetic string which can be connected to a linguistic entity outside that string. What is important is not its meaning, but its arbitrariness.'

Opposing this interpretation, Charles Hockett would speak of the empty morph, a term with which he designates the portion of phonic material morphologically irrelevant, that is to say, a morph that does not correspond to any morpheme of the language. For example, in the form of the Fox language *poon-i-meewa* 'he stops speaking to him', the segment *-i -* lacks semantic content; hence, it should be considered a morph that does not correspond to any morpheme of that language.

This interpretation seems identical to the previous one at first glance. In fact, it also alludes to the appearance in certain words of segments that cannot be assigned a meaning, and, as before, it also allows maintaining the principle of total accountability. However, the conceptual difference among both positions is large: the first one explicitly refuses that morphemes necessarily possess meaning; the interpretation of Hockett, on the other hand, defends the definition of the morpheme as a unit with meaning, because the segments without meaning are not morphemes, but simply, morphs.

However, not all the linguists accept these interpretations, and there are many who prefer to think of words like the ones mentioned in this article as monomorphic, since the analysis of, for example, *remit* into two morphemes can only be based on exclusively formal considerations. However, in approaches that emphasize the meaning and mechanisms that the speaker triggers when using the language, purely formal generalizations may not be of interest.

References

Anderson, Stephen R. Morphological theory. 1988. Linguistics: the Cambridge survey (I. Linguistic theory: foundations), ed. by Frederick J. Newmeyer. Cambridge: Cambridge University Press.

Aronoff, Mark. 1976. Word formation in generative grammar. Cambridge, MA: MIT Press.

Bloomfield, Leonard. 1926. A set of postulates for the science of language. Language 2. 153–64.

Bloomfield, Leonard. 1933. Language. New York: Holt, Rinehart & Winston.

Gleason, Henry A. 1961. An introduction to descriptive linguistics. New York: Holt, Rinehart & Winston.

Harris, Zellig S. 1942. Morpheme alternants in linguistic analysis. Language 18. 169–80.

Harris, Zellig S. 1951. Structural linguistics. London and Chicago: The University of Chicago Press.

Hockett, Charles F. 1947. Problems of morphemic analysis. Language 23. 321–43.

Matthews, Peter H. 1974. Morphology: an introduction to the theory of word structure. Cambridge: Cambridge University Press.

Nida, Eugene A. 1949. Morphology. The descriptive analysis of words. Michigan: The University of Michigan Press.

Robins, Robert H. 1964. General linguistics. An introductory survey. London: Longmans.

JOSÉ CARLOS MARTÍN CAMACHO

See also **Bloomfield, Leonard; Empty Categories; Harris, Zellig Sabbetai; Morpheme; Structuralism**

Endangered Languages

There are some 6,700 known languages in the world, and linguists estimate that 40–90% of them will disappear during this century, a situation generating alarm among linguists and the speakers of languages threatened with disappearance. There is a strong correlation between a language's vitality, or likelihood of being transmitted from generation to generation, and the number of speakers it has. In this respect, the world's languages are unevenly distributed: approximately half the world's population speaks one of just 20 languages, and eight languages (Mandarin, Spanish, English, Bengali, Hindi, Portuguese, Russian, and Japanese) have over 100 million speakers. Ninety-six percent of all languages are spoken by just 4% of the population, and approximately one fourth of all languages have less than 1,000 speakers. In North America alone, fewer than 200 languages remain, although there were certainly hundreds of distinct languages several centuries ago. Today, only a handful of these (e.g. Cree, Dakota, Ojibwa, Navajo)

have a hope of survival, and even their security is doubtful.

In a few cases language loss occurs due to the loss of the speaker population itself. Language change and loss are natural, ongoing processes. Latin, for example, was 'lost' in the sense that its various dialects evolved over time into the modern Romance languages, such as Spanish, French and Italian. Some languages disappear entirely, and it is this absolute language loss that is occurring at an unprecedented rate in modern times. The primary cause is language shift, when speakers cease to speak their own native tongue in favor of the language of what is usually the dominant culture, dominant politically, and/or economically. Such a shift from the heritage to the dominant language can occur over several generations, or even as quickly as over the course of a single generation. In many cases, the oldest generation, the grandparents, speak the heritage language as their first and primary language, the middle generation has some knowledge but uses the dominant language primarily, and the youngest generation has little to no knowledge of the heritage language, and may at most know a few words or phrases. In cases of rapid language shift, however, these changes occur across a single generation.

Language vitality is usually ranked in terms of numbers and generations of speakers. On the one end of the scale are extinct languages that are no longer spoken, and on the other end are viable languages in no current threat of endangerment. In between a number of stages can be recognized. A healthy language with strong vitality is used with a variety of functions and in a range of settings (or domains) of language use. This is true even for bilingual speakers, where one language or linguistic variety is used in personal and home life (e.g. at home, with friends, at social events, etc.) and a different language or variety is used at the workplace, or in more public or official settings. In cases of language attrition, however, the native language is used in increasingly fewer domains with fewer functions.

The causes of language attrition and loss are complex and vary with individual situations. Although significant, the size of the speaker population is not the only reliable predictor of language vitality. Languages with relatively few speakers can be more stable than languages with a larger speaker base. Navajo, a Native American language, provides an example of a language long considered to be stable because of its relatively large number of speakers. It is now considered endangered, due to rapid language shift stemming from a combination of social and economic factors. Two key factors—relative isolation and a strong sense of community identity—can help a language's vitality. The opposite influences—heavy language contact and a weakened sense of ethnic identity (and pride) can increase the likelihood of language shift.

Members of a language community often believe that knowledge of the dominant language is necessary to be successful in the dominant culture, and that knowledge of the heritage language is a hindrance to socioeconomic advancement. This is closely linked to the prestige factor: the heritage language often suffers from a general lack of prestige. There may be no conscious decision to shift from one language to another; rather, the youngest generation learns the dominant language through contact and, frequently, in the schools. Despite the fact that speakers abandon their native tongue for what they deem to be pragmatic and/or prestige factors, the majority of communities that have lost their language subsequently experience this as a deep and regrettable loss.

Native communities and linguists alike are responding to the threat of language endangerment in a number of ways. Language revitalization programs are being implemented in a number of areas; these aim to (re)educate speakers in their heritage language and promote its use. Technical support includes the development of pedagogical materials, dictionaries, the introduction of radio and television broadcasts, and the use of the Internet. Linguists are involved in the recording and documentation of endangered languages and the writing of grammatical descriptions of the endangered languages. They often play a key role in the development of alphabets and literary standards for previously unwritten languages. One crucial and often controversial issue in the development of many literary languages is the selection of the dialect or variety that will serve as the basis for the standard language. Revitalization programs face a number of obstacles, such as a lack of pedagogical materials and trained teachers, and the lack of financial resources to implement any of these.

References

Crystal, David. 2000. Language death. Cambridge: Cambridge University Press.

Dorian, Nancy C. (ed.) 1989. Investigating obsolescence. Cambridge: Cambridge University Press.

Fishman, Joshua. 1991. Reversing language shift: theoretical and empirical foundations of assistance to threatened languages. Clevedon: Multilingual Matters.

Grenoble, Lenore A., and Lindsay J. Whaley (eds.) 1998. Endangered languages: current issues and future prospects. Cambridge: Cambridge University Press.

Krauss, Michael. 1992. The world's languages in crisis. Language 68.

Nettle, Daniel, and Suzanne Romaine. 2000. Vanishing voices. Oxford: Oxford University Press.

Robins, R. H., and E. M. Uhlenbeck (eds.) 1991. Endangered languages. Oxford and New York: Berg.

The most up-to-date information about languages and numbers of speakers can be found in The Ethnologue, ed. by Barbara Grimes, <www.sil.org/ethnologue>

LENORE A. GRENOBLE

See also **Language Death**

English

In 1780, English was a language with fewer than 15 million speakers, most of whom lived in England, Ireland, Scotland, the United States of America, Canada, and the Caribbean. Yet, John Adams foresaw a world where English would play a larger role:

> English is destined to be in the next and succeeding centuries more generally the language of the world than Latin was in the last or French in the present age. The reason for this is obvious, because the increasing population in America, and their universal connection and correspondence with all nations, aided by the influence of England in the world, whether great or small, force their language into general use. (quoted from McCrum et al. 1986:239)

Today, English has an estimated 350 million native and 400 million second-language speakers on all continents. The number of foreign language users easily tops the 400 million mark and is rising as globalization and modern technologies continue to impinge on all areas of modern life worldwide. In terms of political status, English is the national language in *native*, anglophone countries (United States of America, UK, Canada, Australia, New Zealand, Republic of Ireland, Bahamas, etc.), although none of these countries has declared English as its official language. English has (co-)official status in close to 30 countries, most of which had been colonies (e.g. Ghana, Philippines, India, Western Samoa, Costa Rica). It is the dominant language in international institutions (UN, IMF, APEC, EU) and outdoes other languages in international media, companies, conferences, air and maritime traffic, publications, and the web. Sociolinguistically, the norms of English (E) no longer emanate from a single source, even British (BrE) and American English (AmE) are only two of several norm-setting epicenters in a *pluricentric* world of E. National norms also exist in many second-language countries. Foreign language contexts, too, have been claimed to develop norms. It has been said that E is best described in terms of three overlapping circles: native, second, and foreign E. A traditional view, which is adopted here, recognizes a common core across all varieties.

History

As for genetic affiliation, E belongs to the West Germanic (WGmc) branch of Indo-Aryan languages and is related to German, Dutch, and Frisian. Danish, Swedish, and Norwegian are members of the North Germanic (NGmc) family, whose linguistic distance was, at the time of the formation of E, not so large as to inhibit communication—a fact that was relevant during the Old English (OE) period. E lost many WGmc traits during the OE and Middle English (ME) periods and has become a peripheral WGmc member.

Historically, E is the result of the transplantation of the dialects of the Angles, Saxons, Jutes, and Frisians, who invaded England in AD 449, forcing the Celts to the fringes of the islands or to assimilate. The *Anglo-Saxon* or OE phase of E (to c. 1100) witnessed the formation of several kingdoms, three of which (Northumbria, Mercia, and Wessex) gained enough power to create dialects. A national form of E did not develop. The Anglo-Saxons became victims of the Viking invasions from the middle of the eighth century. King Alfred of Wessex was able to defeat them at the end of the ninth century and forced them north of a line from London to Chester (Danelaw). England was partitioned into a NGmc/WGmc north and a WGmc south. Sociolinguistically, the Viking invasions were an obstacle to the formation of a national form of E at first, but after their defeat the sociopolitical and linguistic prestige of Wessex promoted the rise of a quasi-OE standard. Christianization (from the late sixth century) and the political support it had made Latin the language of religion and writing and introduced Latin, Greek, and Hebrew loanwords into E.

The Norman Conquest (Battle of Hastings, 1066) brought about the most significant change in the history of E. For over two centuries, England was under foreign rule. E was demoted to a *low*, rural language, and French and Latin were promoted to the *high* languages in a *triglossic* situation. French gained the status of a national, public language, as well as being the native language of the French aristocracy; Latin was the language of learning, where it was dominant up to the early Modern English period (EModE). There was some bilingualism among the Norman elite and the urban English; but the shift to E occurred when the Normans, facing the loss of their French possessions in the thirteenth century, decided to stay. As E regained status, it was no longer a mere continuation of its OE form. E was a second language for a transitional period, it absorbed thousands of French and Latin expressions, and, under the congruent influence of NGmc and French, lost most inflections and became a fixed word-order language. From 1400, a written standard developed in the Chancery, the oldest national institution; an

educated spoken form developed only from the sixteenth century. Standard BrE and the Received Pronunciation (RP) gradually rose to the position they enjoy today, with the former having greater national and international prestige than the latter, while the former has never ceased to be socially controversial.

As for expansion, E included Cornwall, when Wessex conquered that region during the OE period. The Anglo-Normans founded settlements in coastal, eastern Ireland (The Pale) in the twelfth century, which were soon Gaelicized. The decisive steps forward were made during the Age of Exploration. At the beginning of the seventeenth century, the East India Company transported E to India (1600). In 1607, the foundation of the first American colony, Virginia, transplanted E to America, although some experts argue that whaling had taken E speakers to Newfoundland and Nova Scotia earlier. Slave ships brought E to West Africa and the Caribbean islands. The Bermuda settlement, for instance, was founded in 1609. As for the British Isles, Ireland came under British domination during the reign of Elizabeth I and James I. The Union with Britain in 1801 was a blow to her linguistic independence, leading to the decline of Gaelic, the shift to E, and the development of IrE, which incorporated some Gaelic features. Gaelic has survived in the west and made a recent recovery, partly due to the European Union's language policies. Scotland's southeast had been English from the OE period, but Scotland succeeded in developing a national variety, *Scots*, during ME. The Union of Crowns (1603), the Union of Parliaments (1707), and the defeat of the Highland uprising in the mid-eighteenth century intensified the E influence so much that Scotland was anglicized by the late eighteenth century. Scots declined, but revived as a literary form in the nineteenth century, Gaelic was forced to the Highlands and western islands, Norn to the Shetlands and Orkneys, where it died out in the eighteenth century.

The seventeenth century paved the way for global expansion, and the eighteenth and nineteenth centuries witenessed the most significant geographic and sociopolitical growth. The East India Company, for instance, had been no more than a commercial 'player', with *factories* (trading outposts) in coastal cities. Only by the middle of the eighteenth century did it gain political control over Bengal. It took another 100 years for the subcontinent to become a Crown Colony (1857) and for an Indian elite to emerge that was willing to accept, and later to oppose, E. Similar developments occurred elsewhere in Asia and Africa. Although Anglophone from the beginning, Australia and New Zealand were fully known only by the end of the nineteenth century. The story of the western frontier in America is too well known to require recount-

ing. Again, it was only during the nineteenth century that the United States of America assumed its current extension and it was as late as at the turn of the twentieth century that the United States of America exported their E to the Philippines, Guam (1896), and the Caribbean. AmE was coming into its own and H.L. Mencken (*The American language*, 1919) could declare it to be a 'language', separate from BrE, an allusion to the debates about a 'Federal English' in the late 1770s. The mid-twentieth century witnessed further expansion. World War II took American soldiers to the theaters of war in Asia, Africa, and Europe. Its postwar economic, technological, and cultural power, Britain's enduring role in Europe and the Commonwealth, the problems that former colonies experienced in nation building, and the collapse of communism supported the further expansion of E, which was now fed mostly by the appeal of the United States of America and, logically, AmE.

Today's position of E is, thus, not entirely due to the migratory or imperialist expansion of Great Britain and the United States of America. There was a confluence of factors and it is worth emphasizing that the econo-cultural, technological impact of E was possible only at a particular juncture, as David Crystal explains:

> [But] international language dominance is not solely the result of military might. It may take a militarily powerful nation to establish a language, but it takes an economically powerful one to maintain and expand it. This has always been the case, but it became a particularly critical factor early in the twentieth century.... Any language at the centre of such an explosion of international activity would suddenly have found itself with a global status. And English ... was in the right place at the right time. (1997:8)

Just as postcolonial nations have been willing to retain E, so has the rest of the world where it is a foreign language. Real or perceived benefits attached to E act as powerful incentives and an E language industry guarantees easy access.

The expansion did not remain without consequences for the formal system and lexis of E, norms of use, and attitudes. Up to the seventeenth century, expansion may have amounted to no more than the formation of a single and contiguous British language community. The post-Renaissance period witnessed the formation of more — and distinct — communities: first, the American one (see below), then an (Anglo-) Indian one, and then others. English absorbed *foreign* elements, norms were diluted and others recreated, and attitudes of inferiority gave way to those of pride in one's own E. This development went unnoticed for a long time; it was welcomed in the early United States of America, when the notion of a Federal (American) English was promoted. In the early twentieth century it was, however, perceived as a dangerous disintegration

of the language: 'In our study of the growth of the Empire we forget that this territorial expansion of our language sowed the seeds of disintegration.' (Lloyd James, secretary, BBC language committee 1928; quoted from Leitner 1989:12).

E was a native (ENL), second (ESL), and foreign (EFL) language in widely different contexts, and assumed pluricentric characteristics, with American, British, Australian, Indian, etc., *epicenters*. Malaysia, Singapore, Hong Kong, and other nations face an uneasy choice between condoning the facts, i.e. localization, or endorsing an ideal, i.e. foreign, model. Many countries insist on native-like E but start accepting mixed BrE–AmE forms. Endonormative countries show signs of internal pluricentricity, with national norms coexisting with those of ethnic communities, such as Afro-American E or Australia's Aboriginal E. Although localization became a powerful factor, the disintegration of E was checked by the rise of educated, British-oriented middle classes in many countries during the nineteenth and twentieth centuries. Thus, pluricentricity has not completely diluted family relationships between E varieties. And internationalization is a new unifying factor.

Structure

A formal structural survey can exploit this situation. It will be based on a *common core* and point to areas of variation. Starting with phonology, vowel and consonant systems are quite similar, variation being limited to a few phonemic contrasts between BrE and AmE and, otherwise, being confined to phonetic detail. Scottish E (ScE) and some forms of AmE still distinguish *witch* from *which*. 'Flaps', short, voiced taps of the tip of the tongue against the alveolar ridge, are realizations of /t/ and /d/, e.g. *letter*, in AmE, AusE, New Zealand E (NZE), less so in BrE. Rhoticity differentiates a southern and northern type of E. The former — exemplified by southern BrE, AusE, NZE, and Indian E (IndE) — are nonrhotic and lack postvocalic /r/ in *car*. Rhotic accents, i.e. northern EE , ScE, IrE, AmE, and CanE, retain it. In RP, /l/ has a clear and a dark articulation, cf. *leaf* and *fall*; other varieties, e.g. AusE, may have a single articulation in all positions. /l/-vocalization in postvocalic position is typical of southern EE, especially popular London, and results in a dark vowel, e.g. *school* [skuo]. Vocalization occurs in AusE and NZE. RP has an exceptionally complex and imbalanced vowel system with its 21 mono-, di-, and triphthongs. For instance, *fit/feet*, *foot/food*, *abound/purr* illustrate the only long/short vowel pairs. Words like *pot* and *port* differ qualitatively, not in length. ScE has no phonemic length, length being determined by whether a vowel occurs in open or closed syllables. ScE is, in other words, more systematic than RP. The role of phonetic context on length is worth noting: word boundaries and voiced consonants lengthen vowels (*fee/feed/feet*). Centering diphthongs, e.g. in *pair* or *tour*, mainly occur in nonrhotic accents. An interesting characteristic is that vowels maintain a constant qualitative articulatory difference between them. Thus, all of RP's short vowels are fairly open; all long vowels (and rising diphthongs) start with a rather low vowel. In AusE and NZE, in contrast, short vowels are closer; long ones and rising diphthongs have an open (or back) onset. (Some southern E accents, e.g. London's Cockney, are similar in this respect.) These properties have given rise to speculations about a Second Great Vowel shift in 'southern hemisphere' E, but some experts argue for the conservative maintenance of older features of the East Anglia accent. The reduction of unstressed vowels is another area of variation. RP and many E accents use the *bit* vowel, AusE, NZE and—more advanced RP—have a schwa. Homophony between word pairs like *boxes* and *boxers* is the result. Accents differ most perhaps in lexical phonology. Thus, *either* has /aɪ/ in RP, /i:/ in London's Cockney and genAmE; AusE may have either. From a comparative angle, one should add that E lacks front rounded vowels, which occur in WGmc languages. As for spelling, the AmE and BrE systems differ in a few areas (*center/-re*, *dialog/-ue*, *program/me*, *fueling/-lling*) and more numerously in the spellings of individual words (*check/cheque*). Other varieties follow one or the other system, or attempt to develop a system of their own, as AusE does.

E differs most from contemporary WGmc languages in its paucity of inflectional morphology. Articles lack inflection, nouns have lost most inflectional contrasts (e.g. gender), leaving behind number and the possessive case, which has an analytic variant (*the roof of the car* for *the car's roof*). The genitive can mark an NP (*A friend of y brother's favorite car*) and fulfill a syntactic, phrasal function. As for number, the OE dual has been lost except semantically in *both, either* and *neither*, irregular forms, such as *umlaut* (*goose/geese*), *-en* (*ox/oxen*; *child/children*), and zero (*sheep/sheep*) reflect OE or ME patterns. Verbs have no inflectional contrasts for voice and mood. *Were* is a peripheral remnant of the subjunctive, which is expressed by means of *should* (mainly in BrE) or the bare infinitive (mainly AmE) (e.g. *I suggest he (should) resign from office*). What remains is a third-person singular in the present tense (*he goes* vs. *I/you/they go*), a simple present and past tense (*I like/I liked*), and the progressive, which is marked by *-ing* (in combination with a form of *be*). A striking aspect is a small number of ('strong'), but frequent, irregular verbs. Some verbs have three forms (*go, went, gone*),

others two (*buy, bought*), and still others one (*put*) for the infinitive, simple past and participle, respectively. There are other 'irregularities', e.g. vowel changes (*ta<u>ke</u>, <u>took</u>, ta<u>ken</u>*), remnants of old past tense or participle endings (*tak<u>en</u>, leapt*). As for adverbs, those with initial stress and two syllables have the *-ly* suffix (e.g. *beautifully*). Others are formed analytically (e.g. *in the most friendly fashion* instead of *friendlily). Nonstandard, dialectal varieties, and speech can have unmarked adverbs (e.g. *right*).

There is more inflection in pronouns. Personal and possessive pronouns have a nominative, possessive, and objective case (e.g. *I*, *my/mine*, *me*), and natural gender (in the third-person singular). *Wh*-pronouns have a type of (natural) gender (human/nonhuman, e.g. in *who* and *which*), case (genitive *whose*, objective *whom*). A widely discussed gap is the lack of a second-person singular/plural, which is available in languages like German (*du, ihr*) or French (*tu, vous*). OE had that distinction and, moreover, a dual. Most varieties, especially the standard ones, have lost both, but some have recreated it; consider AmE (*you/you all* or *y'all*, *you guys*) or IrE (*you/youse*), where it is a calque on the Gaelic *tu:sibh* contrast.

The trade-off between inflectional morphology and syntax is a good starting point for a survey of syntax. E is essentially an SVO language; other WGmc languages permit more variation. Thus, German has two standard orders, i.e. 'verb-second' and SOV, in main and dependent clauses, respectively. 'Verb-second' means that any order (SVO, OVS, AVS/O, etc.) is correct as long as V comes second. There is less flexibility in E, but it is not entirely absent, as these examples show: *The young student I like for his sense of humor* (OSV), *Down he came* (ASV), *Here comes the train* (AVS); SOV is impossible, except in verse. Mention must be made of cleft (*It is his sense of humor I like*), pseudocleft (*What I like is his sense of humor*), and existential sentences (*There is an interesting program on TV*). Non-SVO constructions are closely related to the information structure in discourse. Interrogatives manifest subject-auxiliary inversion, which is regular in polar questions (*Can he do that? Wasn't that a great show!*). A note on so-called guided questions, a side issue, is called for, as there is a good deal of variation. They arise when a high rising intonation contour is superimposed on a declarative construction (*So, you are having an ´affair?*) and permit the inference that the proposition contained is assumed to be true. ESL varieties, in contrast, often extend this pattern to ordinary polar questions (*He came home late?*) without that inference. A different pattern has been reported from AusE, NZE, IrE, some northern E dialects, and varieties of AmE: they combine a high rising tone with declarative word order to express social meanings of empathy, friendliness, etc., without the intention of asking a question (*My name is Betty ´Smith*).

Turning to grammatical relations, E contrasts with many languages, including WGmc ones, in ruling out subjectless sentences (G *mich friert* 'I am cold'), although spoken registers like sportscasting have them, if the subject is retrievable. E, however, permits a wide range of semantic roles to occur in subject position, while some of them are more marked in other languages where, as in German, semantic roles are encoded as adverbials. Consider the following examples:

> John can certainly act wonderfully well (Agent)
> This advertising campaign will open up new markets, we hope (Instrument)
> Berlin can be a rather cold city in winter (Location: Place)
> Money is not the problem, really (Object, Theme)

Verbal syntax is another area where E has characteristics of its own. Thus, E has a causative/processual contrast with transitive and intransitive constructions, provided the object of the transitive construction is identical to the subject of the intransitive one:

> John broke the glass
> The glass broke

Ergative constructions contrast with the passive voice, which implies agency without expressing it, while the intransitive-ergative construction attributes some kind of propensity to the subject, a feature that is clearer if manner adverbs occur, as in the following examples:

> The window pane breaks easily
> The line washes well and fast in this washing machine
> Potter's latest book sold extremely well

Two characteristics, i.e. particle verbs and nominal-verbal complexes, deserve mention as very frequent and productive markers of colloquial E. As to the former, several types of such prepositional verbs (*run up* (a hill)), phrasal verbs (*run up* (a bill), (a plan) *takes off*), and phrasal–prepositional verbs (*come up with* (a good idea)) can be identified. Nominalization creates 'particle nouns', which are often marked by stress (*to take* '*off* - '*take-off*). *Take a bath*, *give a smile*, *have a chat* illustrate nominal-verbal complexes that can often be replaced by the verb they contain formally (e.g. *to bathe, to smile, to chat*). There is, one might add, substantial variation with regard to the choice of e.g. *take, have*, etc. E, in contrast to, say, German also has a dynamic *have* (*have a baby*).

Let us close with tense, aspect, and modality. The future is expressed analytically by *will* (*There will be rain in the afternoon*). But there are other ways, e.g. *the chair <u>is going to</u> dip over, be careful*; *the Foreign*

Secretary <u>is about to</u> go to Kuwait; *I'<u>m coming</u>, don't worry*, *We <u>leave for</u> London tomorrow*, which all express subtler shades of future time reference, often merged with modal meanings. A sentence like '*I will come and see you later, John*' is both a promise and a prediction. Apart from being speech act markers, modals can express three types of meaning, i.e. deontic meaning (*I must be gone* = 'I am obliged to'), epistemic meaning (*He can't be there, the light's not on* = from what I know'), and what has been described as 'ability' meaning (*I could swim when I was a small child* = 'I was able'). These meanings interact with negation and question formation in ways that cannot be detailed here (std BrE *You mustn't* (= deontic) *be too strict with the boy* vs. *He *mustn't/cannot* (= epistemic) *be at home yet*), but which suggest that there will be considerable of variation, as there is; in ScE and AusE, for instance, that sentence is grammatical.

Apart from the 'instantaneous present' in sports commentaries (*Beckenbauer kicks the goal*), etc., the simple present has to be paired with the progressive to refer to present time (**Billy goes to school* (*now*) for *Billy <u>is going</u> to school* (*now*)). On its own, it expresses habitualness, etc. The simple past refers to a definite past time, even if it is unexpressed (*Billy went to school*), while the present perfect expresses indefinite past time and notions like current relevance (*I know New York, I have been there* [= indefinite past]; *Shakespeare has written important plays* [= current relevance]. A sentence like **Shakespeare has lived in Stratford*, in contrast, calls for the simple past, as there is no conceivable current relevance. Here, too, there is a great deal of variation. AmE, for instance, has a so-called colloquial simple past; the perfect is more formal (in contrast to, say, German and French, where the opposite would apply). There are constraints on the combination of the perfect with adverbs, such as *yet*, *already*. Once again, varieties differ a great deal. Std BrE, for instance, disallows **Did John come in yet*; AmE, AusE, BrE do not.

Lexis is especially interesting from a pluricentric perspective. The quite homogeneous, original WGmc word stock has been enriched by tens of thousands of expressions that E has absorbed from languages it has been in contact with. As a result, such words have been integrated; others have been lost or specialized to new meanings, uses, or stylistic patterns. Contact with Celtic and NGmc dialects provided the first nonnative influence. NGmc introduced some frequent words like *law*, *fellow*, *they*, and *both*. Some of them became referentially distinct (*skirt–shirt*), others dialectal (*kirk*, ScE). *Brogue*, *whisk(e)y*, and *Tory* are Celtic in origin. Latin has influenced E from as early as OE, even prior to the WGmc tribes settling in England, as *street* and *kitchen* suggest. Late OE loans from Latin are *sign*,

clerk, and *demon* (the latter two of Greek origin). French has exerted a pervasive and ongoing, if decreasing, influence toward the nineteenth century. With Romance words being so frequent, they barely require illustration, except to mention that Romance (and Greek) introduced a differentiation of style. Gmc words tend to be common and/or informal, Romance ones formal and/or technical: consider *begin-commence*. Romance also increased the potential of E word-formation patterns that derived from Wgmc, e.g. *kingly*, *royal*, and *regal*, and introduced new ones, e.g. in *-ify* (*beautify*), *-ate* (*nominate*), or *describe-descriptive-description*. The latter cannot be analyzed in terms of morpheme sequencing like *king+<u>dom</u>*. As a result of borrowing, the Gmc word stock is now a low 30% and the Romance one is 50%.

We now illustrate the absorption of expressions from languages that E in its standard E (ENL) and in ESLs varieties has come in contact with during its global expansion: American-Indian (*caucus* fr. Algonquian; *hickory*), Arabic (*zero*; *assassin*), Hindi (*sari*; *bandh call* 'call for a strike'), Telugu (*bandicoot*), Australian Aboriginal languages (*boomerang*; *cooee* 'type of call'), Malaysian/Singaporean E (*la* 'emphatic particle, fr. Chinese), African languages (*baobab*), Japanese (*rickshaw*; *karaoke*), German (*angst*; *-friendly*, see 'user-friendly), Malay (*jubah* 'in Islamic culture, a kind of clothing'), Spanish (*provinciana/o* 'woman/man from provinces, considered unsophisticated', current in Philippine E). Add to this loan translations, e.g. *acupressure* [< Jap.], *walkabout* ('walk in public', < Australian Aboriginal languages), or the productive use of some patterns in e.g. *crash-friendly* (cars).

The most important feature of all is the lexical creativity of new varieties. The impact of AmE on science, technology, the media, sports, informal E, and slang is too well known to require illustration. To mention a few older words, e.g. *to trade* 'to shop'; *to ship* 'to send'; *log-rolling*; *on the fence*, some of which echo the moving frontier society. Generally speaking, AmE is more informal and its dominant form is defined as *general* AmE, in contrast to the more elitist definition of standard BrE, especially RP. Well-known expressions in AusE are *mate* 'pal, friend', *to barrack for* 'to support (e.g. a sports team)', *bushie* 'someone who lives in the bush', in IndE *carlifter* 'car thief', *handphone* in Asian E. Localized idiomatic expressions, such as *We must meet again* (AmE), *Can I leave now*, or *Have I your permission to leave, Sir* (IndE, PakE 'it's been nice talking to you'), are highly visible and invite misunderstanding, as do words like Australian AborE *women's business* 'knowledge confined to (eligible) women'. Others such as *to jump the dragon gate* in Chinese E may not be understood at all.

Many expressions remain confined to local varieties, hence the pluricentricity of E; but many have become common and have made E a lexically rich language, as the *Oxford English dictionary* and *Webster's third international dictionary* and national dictionaries in Australia, India, etc., demonstrate.

With respect to word formation, three processes will be mentioned. The first is conversion or zero derivation, the most productive pattern today, which permits words to change word class membership without a formal marker (*chair—to chair*). Often, words occur in several word classes (*right* as a preposition, noun, verb, adjective, or adverb). Similar to conversion is the derivation of noun–verb pairs by means of a stress shift, as in '*import—to im'port*, '*take-off—to take 'off*. The second is the deletion of a preposition in prepositional verbs, which results in an ordinary transitive verb. Thus, *to appeal against a decision* becomes *appeal a decision*. The inverse process inserts a preposition (*to meet with*). The third is clipping (*bod < body*) and suffixation in *-i(e)* and *-o* (*pollie/y* < 'politician', *garbo* < 'garbageman'). Phonological adaptations may be necessary, e.g. *mossies* (< *mosquitoes*) with a /z/, like *Aussies*, instead of /s/. An interesting aspect of clipping is the formation of first names to express solidarity, acquaintance, etc.; consider *Kez* (< Kerry), *Baz* (< Barry) These processes are common in E, but their application may result in words that are specific to some variety of E. ESL varieties do not seem to add processes not found in native E either. Hybridization may be more frequent (IndE *bandh call* 'call for a strike'), although it also occurs in ENL context when a borrowed word may be difficult to understand (e.g. *koala bear* in AusE).

Formation of 'Englishes'

This *tour de force* has shown that the transplantation of dialects into new *habitats*, dialect contact, and contact with local languages have been the major factors that have shaped the development, texture, and status of E. Diversification has been the predominant outcome, the loss of a variety being rare. But it has happened, as the decline of Scots shows. Centrifugal and centripetal forces have always been active. Two themes must be added to this account, viz. the dating of important BrE and AmE bipolarity and English-based contact languages, especially pidgins and creoles.

The 'great divide' between the two main 'epicenters' of E, viz. BrE and AmE, has been a slow one, although signs of separation had been seen early. With respect to phonology, the divide has been dated at around 1750. The loss of postvocalic /r/ and of secondary stress in suffixes ('*secretary* vs. AmE '*secre,tary*) have remained confined to RP; others con-sider the unrounding of the *lot*-vowel, yod-dropping in *tune*, to genAmE. The dialect divide is difficult to date, but Webster's *American dictionary of the English language* (1828) is probably as good a marker as any. It marks differences in lexis, spelling, lexico-grammar, and grammar. Apart from the national element in American debates, a strong, shared, prescriptive tradition was visible during the late eighteenth and nineteenth centuries and counteracted separation. Bishop Lowth's, Murray's, or Cobbett's grammars, etc., catered to markets on both sides of the Atlantic. Prescriptivism continued, for instance, with Henry Alford's *The Queen's English* (1864), H.W. and F.G. Fowler's *The King's English* (1906), and H.W. Fowler's *A dictionary of modern English usage* (1926). Although they explicitly denied writing for an American market, which they perceived as different, Fowlers' books were used there. Debates about *Webster's third international dictionary* and English curricula in the United States of America show the strength of prescriptivism today. It is slowly giving way to descriptive approaches. The animosity of the British against AmE, which became a public topic at the end of the eighteenth century, has survived all intellectual shifts and extended its reach to practically all varieties of E of BrE descent.

Although they are considered distinct languages, E-based pidgins and creoles must be mentioned as they show, among others, the power of the language in intercultural situations. Pidgins typically emerged in colonization scenarios in Africa (e.g. Krio), the Caribbean (e.g. Jamaican Creole), southern states of the United States of America (e.g. Gullah), Australia (e.g. Kriol), India (e.g. Butler English), and the Pacific (e.g. Tok Pisin in New Guinea). Controversies about what factors gave rise to them, the formative role of superstrate E vis-à-vis local, substrate languages, the diffusion of pidgins through maritime traffic, and the role of universal processes of language development remain unresolved. The fact remains that pidgins and creoles are similar, whether one looks at North America, Africa, or the Pacific, where genetically and typologically different languages are used. The availability of E and upward social mobility draws them into the realm of mainstream E and they tend to become the (socially) *low*, nonstandard, end of a continuum that goes *up* to (national or international) standards of E. Ethnic forms, e.g. Afro-American E or Aboriginal E, are in a middle range. Speakers can exploit the continuum for communicative purposes, but it would be wrong to assume a simple 'upward' orientation, with the highest variety having the highest prestige. The 'high' variety may be merely used in outside communication or selected domains. Contact varieties often symbolize ethnic identities and show

verbal repertoires that differ significantly from mainstream E, such as Afro-American E or Jamaican Creole. The dialect continuum may be less real, sociolinguistically, than a formal analysis suggests.

John Adams's prediction of a global role of E has come true. But it is not the E that he may have had in mind. It is not bicentric, it is pluricentric, subject to conflicting forces that emphasize local and global trends. The future is hard to predict, but it seems unlikely that it will encourage disintegration.

References

Bailey, Richard W. 1991. Images of English. A cultural history of the language. Cambridge: Cambridge University Press.

Bailey, Richard W., and Manfred Görlach (eds.) 1983. English as a world language. Ann Arbor: University of Michigan Press.

Brutt-Griffler, Janina. 2002. World English. A study of its development. Clevedon: Multilingual Matters.

Burchfield, Robert. 1985. The English language. Oxford, New York: Oxford University Press.

Conrad, Robert, Joshua Fishman, and Robert Cooper. 1977. The spread of English. Rowley, MA: Newbury House.

Crystal, David. 1997. English as a global language. Cambridge: Cambridge University Press.

Dillard, J.L. 1992. A history of American English. London: Longman.

Fisher, John. 1996. The emergence of Standard English. Lexington, KY: The University Press of Kentucky.

Görlach, Manfred. 1991. Englishes: studies in varieties of English. Amsterdam: Benjamins.

Görlach, Manfred. 1999. English in nineteenth century England. An introduction. Cambridge: Cambridge University Press.

Graustein, Gottfried, and Gerhard Leitner (eds.) 1989. Reference grammars and modern linguistic theory. Tübingen: Niemeyer.

Kachru, Braj. 1986. The alchemy of English. Oxford: Pergamon Press.

Leitner, Gerhard. 1982. The consolidation of "educated southern English" as a model in the early 20th century. IRAL 10(2). 91–107.

Leitner, Gerhard. 1986. The English reference grammar. Language and linguistics, writers and readers. Tübingen: Niemeyer.

Leitner, Gerhard. 1989. BBC English und Englisch lernen mit der BBC. München: Langenscheidt.

Leitner, Gerhard (ed.) 1991. English traditional grammars: an international perspective. Amsterdam: Benjamins.

Leitner, Gerhard. 1992. Pluricentric English. Pluricentric languages, ed. by Michael Clyne, 178–237. Berlin: Mouton de Gruyter.

Maurais, Jacques, and Michael Morris (eds.) 2003. Languages in a globalising world. Cambridge: Cambridge University Press.

McArthur, Tom (ed.) 1992. The Oxford companion to the English language. Oxford: Oxford University Press.

McArthur, Tom. 1998. The English languages. Cambridge: Cambridge University Press.

McCrum, Robert, William Cran, and Robert MacNeil. 1986. The story of English. London: Faber & Faber.

Milroy, John, and Lesley Milroy. 1985. Authority in language. London: Routledge and Kegan Paul.

Schneider, Edgar (ed.) 1996. Focus on the USA. Amsterdam: Benjamins.

Trudgill, Peter, and Jean Hannah. 1994. International English. A guide to the varieties of standard English. London: Arnold.

Wells, John. 1982. Accents of English, 3 vols. Cambridge: Cambridge University Press.

GERHARD LEITNER

See also **Celtic Languages; Indo-European 2: Germanic Languages; Old English**

Epenthesis and Syncope

Syncope (from Greek 'cutting off') and epenthesis (from Greek 'insertion') are cases of the more general phenomena of the deletion and insertion of segments, respectively. Initially utilized in the description of historical phonological change, use of the terms has been extended to characterize synchronic alternations as well. Most frequently, 'syncope' refers to the dropping of a word-internal unstressed vowel, as in Latin CALIDU(M) 'hot', NOBILE 'noble', VIRIDE(M) 'green', which became Spanish *caldo*, *noble*, *verde*, respectively. Expanded usage of the term 'syncope' covers dropping of a final vowel (also called 'apocope') as in Latin BLANCU(M) 'white', MURU(M) 'wall' becoming French *blanc*, *mur*, respectively.

Further extension to sounds (both vowels and consonants) in word-initial syllables (also known as 'aphaerisis') is less common, no doubt matching the lesser frequency of the phenomenon itself given the usual salience of initial syllables. However, it does occur, as reflected in the loss of initial /k/ in such English words as *knife* and *knee*, the initial vowels in *bishop* (from Latin *episcopus*), *possum* (from Algonquian *opossum*), *special* (from Old French *especial*), or the disappearance of initial /h/ in English dialects (Cockney being the typical example).

'Epenthesis', too, designates a variety of phenomena involving the insertion of both vowels and consonants. Consonantal epenthesis (where the appropriate

conditions are created by the previous syncope of Latin E) is illustrated in (1):

(1) | Latin | Old French | |
|---|---|---|
| CIN(E)REM | *cendre* | 'cinder' |
| *ESS(E)RE | *estre* | 'to be' |
| MOL(E)RE | *moldre* | 'to grind' |
| NUM(E)RU(M) | *nombre* | 'number' |

Note that, historically, English *empty* and *thunder* show epenthetic /p/ and /d/, respectively. Moreover, the English words (French loans all) *humble* — *humility*, *cinder* — *incinerate*, *number* — *enumerate*, etc., show both the remnants of historical epenthesis at the Old French stage and a minor synchronic alternation. This Old French epenthesis underlies a synchronic alternation occurring in a number of irregular verbs in Modern French: *connaître* — *connaissons* 'to know', *coudre* — *cousons* 'to sew', *craindre* — *craignons* 'to fear', *falloir* — *faudra* 'to be necessary', *tenir* — *tiendra* 'to hold', *valoir* — *vaudra* 'to be worth', *venir* — *viendra* 'to come', among others.

Consonantal epenthesis has been investigated using both phonological considerations and phonetic approaches. In phonological (syllable structure) terms, epenthesis often reinforces the onset of the second syllable. Thus, in *cendre*, earlier /n-r/ becomes /n-dr/ where syllable-initial /dr/ provides a stronger onset. In phonetic terms, the relative timing of nasalization or fricative cessation relative to the following segment results in the insertion of a transitional or epenthetic consonant.

'Epenthesis' can also refer to the insertion of a vowel in various positions in the word. Initial vocalic epenthesis (also known as 'prothesis') is seen in Latin SCHOLA 'school', SCRIBET 'he writes', STELLA 'star' becoming Spanish *escuela*, *escribe*, *estrella*. Word-internal vocalic epenthesis (also called 'anaptyxis' or 'svarabhakti', the latter from Sanskrit where the process is common) occurs variably in such nonstandard English pronunciations as *athalete* [æθəlijt], *artheritis* [aɹθəɹajDɪs], *fillum* [fɪləm] (standard *film*), while final epenthesis, normally of a consonant and much rarer, is seen in such English forms as *amongst*, *whilst* and nonstandard *acrosst*, *heightth* (with analogical pressures from *length*, *width*, and *depth* being the obvious source in the latter case). Historically, the final stops in *pheasant* and *sound* (< Old French *fesan* and *son*, respectively) are also epenthetic. As the foregoing discussion indicates, the addition or deletion of segments can occur in a variety of positions in the word, with a correspondingly complex set of terms reflecting this behavior. Not all scholars use all terms, moreover, and their use is not always consistent. The practice

described here — syncope for loss of a word-internal unstressed vowel; epenthesis as a more general term for vowel or consonant insertion, again word-internally — seem to reflect the most widespread usage (see Lass 1984:183–8 for a thorough survey).

Both syncope and epenthesis (in either their more limited or wider senses) are interesting because of the general theoretical issues they raise. Both are linked to phonotactic considerations — the constraints on sequences of segments or on syllable structure. Consonantal epenthesis often results in the addition of syllable-initial sounds and strengthens, as a consequence, syllable onsets (cf. *cendre* above). Vocalic epenthesis has the effect of separating consonants whose proximity violates the restrictions on consonantal combinations and leads to an improved syllable structure. Thus, in *athlete (athalete)*, the rare /θl/ sequence is interrupted by the insertion of schwa and a VC.CVC syllable structure (where V = vowel and C = consonant) becomes the more acceptable V.CV.CVC. Syncope, in addition to phonotactic considerations, is also sensitive to accentual and rhythmic patterns (accented vowels and vowels in initial syllables are rarely involved). Speech rate and word frequency also play a role: deletion of vowels is fostered by rapid speech, as we see in pronunciations such as /ptéjDow/ (or even *taters*) for *potato*, and the greater frequency of *memory* (> [mɛmɹij]) vs. *mammary* facilitates deletion of the penultimate vowel in the former. Finally, if the operation of syncope were to create a prohibited sequence, it is frequently the case that deletion is blocked or delayed. Consider the Latin forms in (2) where the vowel preceding stress would normally be expected to delete:

(2) | Latin | Old French | |
|---|---|---|
| CAPRIFOLIU | chevrefeuil | 'honeysuckle' |
| *PUTRITURA | porrëure | (Mod. French *pourriture* 'rottenness') |
| QUADRIFURCU | carreforc | (Mod. French *carrefour* 'crossroads') |

In each case, the highlighted vowel in Latin is retained as schwa in Old French because deletion would have produced the unacceptable consonant clusters PRF, TRT, and DRF, respectively. (Later simplification of the consonant groups obscures the original conditions.)

Syncope and epenthesis, as characteristic of deletion and insertion in general, are directly involved with a wide variety of central phonological issues. They are relevant to considerations of syllable structure and phonotactic constraints in general, of the rhythmic properties of speech, rule naturalness, and the use of phonetic explanations in phonology, and no doubt others. As such, the two processes merit wide-ranging and detailed attention.

295

References

Lass, Roger. 1984. Phonology. Cambridge: Cambridge University Press.

Murray, Robert. 1989. On epenthesis. Folia Linguistica XIII (3–4). 293–316.

Trask, Larry. 1996. Historical linguistics. London: Arnold.

Walker, Douglas. 1978. Epenthesis in Old French. Canadian Journal of Linguistics 23(1). 66–83.

DOUGLAS C. WALKER

See also **Phonetics; Phonology; Syllable: Structure**

Eskimo-Aleut

The Eskimo-Aleut language family is spoken from the Chukotkan Peninsula on the Siberian coast in the West to Greenland in the East. It includes two major branches, Aleut and Eskimo. Eskimo consists of at least two further subgroups, Yupik and Inuit, and there is some question as to the proper classification of Sirenikski, which may be a Yupik language or may constitute a third branch of Eskimo. The terms 'Eskimo' and 'Aleut' both have unclear origins, and they are not universally in use within the respective areas. Thus, it is commonly held that 'Eskimo' was derived from a derogatory Algonquian term meaning 'eaters of raw meat'; the negative association has led to the general replacement of 'Eskimo' with *Inuit*, the native term for 'people', in most of Canada, and there are various designations for the language, including *Inuttun, Inuktitut*, and others. The term 'Aleut' was bestowed on the people native to the Aleutian Islands and their language by Russians in the eighteenth century; the native term in use today for self-designation is *Unangan* or *Unangas* for the people and *Unangam Tunuu* for the language. Because there is still no other general term to describe all of the languages and dialects encompassed by the term 'Eskimo', and for reasons of linguistic tradition, both terms are still commonly used within the field of Eskimo-Aleut linguistics.

It is generally accepted that the Eskimo and Aleut people were part of a relatively late migration from Asia across the Bering Land Bridge, sometime between 4,000 and 6,000 years ago. The development and differentiation of Eskimo-Aleut are believed to have taken place in Alaska because of the linguistic diversity found on the Alaskan side. From Alaska, there was continued migration down to the Aleutians and eastward to Greenland. The split between Aleut and Eskimo is thought to have taken place around this time, although this figure is not undisputed, and there is archeological and skeletal evidence to suggest that the divergence between Aleut and Eskimo may be at least 9,000 years old. Various attempts have been made to link Eskimo-Aleut with other language families on the Asian continent, and thus trace the history of the language family even further back in time. While there is little solid linguistic evidence of a genetic relationship between Eskimo-Aleut and other language families, there are strong suggestions of very early contact, particularly with Uralic (for a thorough discussion of possible linguistic affinities and contact, see Fortescue 1998).

Within the Eskimo branch, the split between the Yupik and Inuit branches must have taken place about 2,000 years ago, and there is good linguistic evidence (e.g. shared phonological and prosodic features between Yupik and neighboring Inuit speakers in areas now no longer predominantly Yupik) for positing the prehistoric (i.e. prior to European contact in the eighteenth century) presence of a continuum of Yupik speakers around the Bering Strait. From their original homeland in Alaska around the Seward Peninsula, Yupik speakers gradually occupied southwestern Alaska and, moving westward across the Bering Strait, reoccupied the Chukchi Peninsula in Siberia. Central Siberian Yupik and neighboring Chukchi have gradually displaced Sirenikski, and the latter is now extinct. If Sirenikski is a separate branch of Eskimo, then it may have split off about 2,500 years ago and its origins may be on the Chukchi Peninsula. About 1,000 years ago, there was another eastward migration out of Alaska; known as the Thule culture, this very rapid migration is associated with the spread of the present-day Inuit language. The present dialect differentiation is possibly as recent as the past 500 years (see Dorais 1996). There is also evidence for the spread of Inuit speakers into Yupik-speaking territory in the Seward Peninsula in the past two centuries.

There are certain features that, taken together, are particularly characteristic of the Eskimo-Aleut language family. Thus:

- All languages have three basic vowels (i, u, a), which are derived from an original four-vowel system (i, u, a, and schwa, represented by [e] in Yupik, which maintains the four-vowel

distinction; [e] will be used to represent schwa in this discussion).

- All of the languages are almost exclusively suffixing, the only exception being an anaphoric prefix *ta-* on demonstrative stems (e.g. pan-Eskimo *una*, *tauna* 'this one').

- All are polysynthetic: to a greater or lesser degree within the language family, they allow for very complex verb structures that encode meanings for which other languages need whole sentences. This is particularly developed in Canada and Greenland; e.g. Holman Island Inuinnaqtun (example from Lowe, in Tersis and Therrien 2000:167):

iqaluk-hiu-riaqtu-qati-gi-tqiktaqpak-kalua-ramiung
fish-hunt-go.to.do-partner-have-several.times-even.though-he:him
'even though he went fishing several times with him'

- All are clause-chaining: sentences typically consist of a series of dependent clauses headed by an independent clause. In the following relatively simple example from West Greenlandic, there is one indicative clause and two subordinated participial clauses; clause chains can be quite extensive (example from Fortescue 1984:39):

irn-i   *qajartur-tuq*
son-his.own   be.out.in.kayak-his

qinnguar-paa
see.through. binoculars-he:him

natsirsu-up   *sursuk-kaa*
hooded.seal attack-it/his
'he$_i$ saw his son$_j$ through his binoculars being attacked by a hooded seal while in his$_j$ kayak'

- All have SOV (subject–object–verb) word order to a more or less fixed degree (Aleut has essentially fixed word order). Note that in the previous example, the object clause *irniq qajarturtuq* precedes the verb, and the subject of each of the subordinated clauses precedes its respective verb.

- All have, or originally had, a largely ergative–absolutive case-marking system (with extreme modifications in Aleut), where the subject of intransitive verbs receives the same marking as the object of transitive verbs, namely absolutive case, while the subject of transitive verbs receives a different marking: ergative case.

The Eskimo languages are much more closely related to each other than to Aleut. In addition to the common features listed for Eskimo-Aleut, they share

- certain restrictions on syllable and other phonological structures,
- up to four or five nongrammatical cases, in addition to ergative and absolutive cases (locative, modalis, ablative, allative, vialis, and equalis),
- transitive and antipassive structures, whose choice appears to be partially determined by definiteness, as in the Central Alaskan Yup'ik examples below:

Angute-m   *neqa*   *ner-aa*
man-ERG   fish-ABS   eat-he:it
'The man eats **the** fish'

Angun   *neq-mek*   *ner'-uq*
man-ABS   fish   eat-he
'The man eats **a** fish'

In the descriptions of the languages, there are also certain features that are commonly used to differentiate the individual languages or dialects. Below, Aleut, Yupik, and Inuit are described in terms of their major languages or dialects, the major linguistic features that differentiate them from each other, and their current status and outlook.

Aleut is a language with two clearly distinct, extant dialects, and at least two other dialects historically attested. Eastern Aleut is spoken east of Amukta Island to the Alaskan Peninsula, as well as on the Pribilof Islands. Atkan, also variously called Western or Central Aleut, is today spoken on Atka Island, and a version of it is spoken on Bering Island. A third dialect, Attuan, is essentially extinct, and it had close affinities with Atkan; a mixed language known as Copper Island Aleut, consisting of Attuan stems and Russian inflection, is still spoken on Bering Island. In very early descriptions, there was apparently a separate dialect spoken on the Rat Islands in between Attu and Atka, although almost no extant information is available on this form of Aleut. Characteristic of Aleut are

- phonological features such as its lack of a [p], and its aspirated nasals,
- consonant clusters that differ from those in Eskimo in their distribution (e.g. they are permitted word-initially, as in *sngaxsix* 'to dream'), in the combinations of consonants possible (e.g. velar–uvular fricatives), and in their complexity (allowing up to three consonants, as in Eastern Aleut *ixchxingin* 'his neck'),
- use of independent pronouns, as opposed to the complex verb structures of Eskimo languages:

Aleut   *txin*   *achix-ku-qing*
              you   teach-presently-I
              'I am teaching you'

Greenlandic   *ilinniar-tip-pakkit*
learn-cause-I:you
'I am teaching you'

- its typologically unusual agreement system, in which ergative case marking is only used if a transitive object or an object of posession is not overtly expressed:

  Tayaĝu-x̂   qa-x̂   qa-ku-x̂
  man   fish   eat-presently
  'The man eats the/a fish'

  Tayaĝu-m   qa-kuu
  man-ERG   eat-presently:he:it
  'The man eats it'

Modern Aleut also has a large proportion of Russian loanwords, the result of extensive contact with Russian traders and colonizers in the eighteenth and nineteenth centuries. As a result of early decimation of the people and later suppression of the language in schools, the language is severely endangered today, with at most 150 speakers in the Aleutian Islands, the Pribilof Islands, and Anchorage, and perhaps eight to ten on Bering Island. With the exception of Atkan, the speakers are elderly.

The Yupik languages include Naukanski, spoken around East Cape on the Chukchi Peninsula; Central Siberian Yupik, also spoken on the Chukchi Peninsula as well as on St. Lawrence island; Central Alaskan Yup'ik, spoken from Norton Sound to Bristol Bay in Alaska; and Sugpiaq, also known as Sugcestun or Alutiiq, spoken around Prince William Sound, the Alaskan Peninsula, Kodiak and Afognak Islands, and the tip of the Kenai Peninsula. Yupik languages are characterized by

- their retention of a fourth vowel that presumably stems from the ancestral (proto) Eskimo language (cf. Proto-Eskimo **neqe* became Central Alaskan Yup'ik *neqa* 'food', Iñupiaq *niqi* 'food')
- more or less complex effects of intonational stress; in stressed syllables, for example, the vowel is often lengthened (for more on Yupik prosody, see Krauss 1985).

There are some nonnegligible syntactic differences between the languages, although these have not yet been well described. Siberian Yupik languages have a number of English loan words, from contact with nineteenth-century whalers, and Alaskan Yupik languages have a large number of Russian loans from eighteenth to nineteenth century Russian colonization, as well as twentieth-century English loans. Most Yupik languages are severely endangered today, with numbers of speakers ranging from 70 (Naukanski) to 1,300 (Central Siberian Yupik). The notable exception is Central Alaskan Yup'ik, with about 10,000 speakers

(and on the Kenai Peninsula these include children), and with immersion programs in the schools and active production of learning materials.

Sirenikski is considered related to the Yupik languages; however, it has also been regarded as a separate branch of Eskimo. It is seen as an important link to Proto-Eskimo because of particularly conservative features such as a retention of consonants between vowels, which were lost in all other Eskimo-Aleut languages (e.g. Proto-Eskimo **ataʀuciʀ*, Sirenikski *ateɡesegh*, Central Alaskan Yup'ik *atauciq,* Iñupiaq *atausiq* 'one'). It has, however, undergone sound changes quite different from other Eskimo languages. For example, in all nonstressed (essentially noninitial) syllables (as in the example given above) the vowel changed to schwa. Unfortunately, it was first discovered and described at the end of the nineteenth century, when it was already highly moribund; the last speaker died in the year 2000.

Inuit is generally described as a language with four distinct dialect groups, each of which have their own recognizable subdialects; these groups include Alaskan Iñupiaq, with four major subdialects spoken from the Seward Peninsula and north; Western Canadian Inuit, with four major subdialects and spoken over a vast area of Central Arctic Canada from MacKenzie Coast to Repulse Bay; Eastern Canadian Inuktitut, with six major subdialects, spoken in Baffin Island, Arctic Québec, and Labrador; and Greenlandic Kalaallisut, with three major subdialects and spoken in Greenland (there is also a sizeable population of speakers in Denmark). Characteristics of Inuit are

- lack of intonational stress as compared with Yupik (with noticeable difference in intonational patterns between the dialect groups),
- loss of the fourth vowel, with various important phonological traces,
- various degrees of consonant and vowel cluster simplifications (cf. Iñupiaq *aglaun* 'pen', Greenlandic *allaat* 'pen', in which *gl became ll* and *au turned into aa*), and
- a tendency to merge parts of the tense system most important in narration, with most extensive merging in Alaska and least in Greenland.

In Alaska and Canada, large numbers of English loans are found; in Eastern Canada, there are also a few loans from French and German through the influence of missionaries. In Greenland, on the other hand, loans are predominantly from Danish. The status and viability of the language are quite different in the different regions. In Alaska and Western Canada, the language is severely endangered, with only about 3,000 speakers, almost none of whom are children. In Eastern Canada, there are about 20,000 speakers, but there is widespread bilingualism in almost all age-groups and a growing tendency for

English to replace Inuktitut. Active efforts are under way to reverse this process, including the encouragement of Inuktitut programs in schools. In Greenland, however, over 95% of the native population of some 50,000 are speakers of Kalaallisut, and the language is thriving.

References

Bergsland, Knut. 1997. Aleut Grammar/ Unangam Tunuganaan Achixaasix, Research Paper Number 10. Fairbanks: Alaska Native Language Center.

de Reuse, W.J. 1994. Siberian Yupik Eskimo. The language and its contacts with chukchii. Salt Lake City: University of Utah Press.

Dorais, Louis-Jacques. 1996. La Parole Inuit: Langue, Culture et Société dans l'Arctique Nord-Américain (The Inuit Word: language, culture, and society in the North American Arctic), Paris: Peeters.

Fortescue, Michael. 1984. West Greenlandic. London: Chroom Helm.

Fortescue, Michael. 1998. Language relations across Bering Strait: reappraising the archaeological and linguistic evidence. London and New York: Cassell.

Jacobson, Steven A. 1995. A practical grammar of the Central Alaskan Yup'ik Eskimo language. Fairbanks: University of Alaska, Alaska Native Language Center.

Kleinschmidt, S. 1851. Grammatik der Grönländischen Sprache (Grammar of the Greenlandic Language). Berlin: G. Reimer.

Krauss, Michael (ed.) 1985. Yupik Eskimo Prosodic systems: descriptive and comparative studies. Alaska Native Language Center Research Papers 7. Fairbanks: ANLC.

MacLean, Edna Ahgeak. 1986. North Slope Iñupiaq grammar: first year, 3rd edition revised. Fairbanks: ANLC.

Menovshchikov, Georgiy Alekseevich. Yazik Sirenikskikh Eskimosov: Fonetika, Ocherk Morfologii, Teksty i Slovar (The language of the Sireniki Eskimos: phonetics, morphological sketch, texts, and dictionary). Moscow and Leningrad: Isdatelstvo Nauk.

Tersis, Nicole, and Michèle Therrien (eds.) 2000. Les Langues Eskaléoutes: Sibérie, Alaska, Canada, Groënland (The Eskimo-Aleut Languages: Siberia, Alaska, Canada, Greenland). Paris: CNRS.

Veniaminov, Ioann. 1846. Opyt grammatiki aleutsko-lis'evskago yazyka (An essay on the grammar of the Fox dialect of the Aleut language). St. Petersburg: Academia Nauk.

ANNA BERGE

See also **Canada**

Ethnicity and Language

The words 'ethnicity' and 'ethnic' were in common use at the end of the 20th century in contexts so widely disparate that no common definition will suffice to unite the variety of meanings. We speak of 'ethnic studies', 'ethnic groups' and 'ethnic neighborhoods', 'ethnic and racial groups', of 'ethnic revivals', and 'ethnic cleansing'—a euphemism for genocide that came into use in the 1990s when certain countries tried to drive minority 'ethnicities' from their territory through terror and murder. In the 2000 census of the United States, although the Census Bureau of the United States used traditional categories of 'race' such as White, Black, Asian, or Pacific Islander, American Indian or Alaska Native, Other, and Multiracial in gathering census data, any of these might equally well be regarded as 'ethnic' classifications in today's language usage. It is unclear where ethnicity leaves off and race begins. 'Hispanic', in the 2000 census, was stipulated by the Census Bureau as 'may be of any race'—which indicates the extent to which 'race' and 'ethnicity' overlap in contemporary discourse.

Ultimately, as we use the word today, ethnicity is not a matter of strict definition. It is a matter of identity: you are what you say you are and what other people think you are. We find it convenient in certain contexts to use the phrase 'ethnic group' for a wide variety of minorities in America and other countries: Irish Americans, Italian Americans, Asian Americans, immigrant South Asians and their children, Native Americans, African Americans, Roma (Gypsies) in the many countries where they live, French Canadians in Canada and Maine, Kurds in Turkey, Basques in Spain and France. In recent years, advocates of the interests of gay and deaf communities in America have argued for the benefits of using the term 'ethnic' in referring to these groups. 'Ethnicity' in current usage is so elastic and so convenient a term of reference that it may perhaps most usefully be defined simply as the 'Other'—a member of any minority group in a country who retains or is thought to retain the distinctive characteristics of that group.

Otherness links ethnicity and language. Language always has a setting. If this setting is the majority of a country's population, then the language of that majority will be the 'language of the country'. If this setting is an ethnic minority of a country's population, then the language of that ethnicity will either be a language different from the language of the country or a dialect

of the majority language that nevertheless differs significantly from it. Language does not exist apart from ethnicity; it is a visible badge of ethnicity. A different ethnicity therefore often implies a different language, or at least a different variant of a majority language. (We sometimes refer to ethnic dialects as 'ethnolects', modeled after the pattern of 'dialect' and 'idiolect', the speech of a single individual.)

From the end of the Civil War to the end of World War I (1865–1918), the population of the United States grew rapidly through immigration. There were prejudices and laws against immigration of members of 'races' thought to be inferior or 'incompatible' with the majority population of America—against Asians and Africans for example—but for others, gaining admittance to America was not difficult. This led to massive immigration from Ireland, Germany, Italy, Poland, Russia, Scotland, Sweden, England, and other predominantly European countries. Jews, who had faced discrimination and pogroms in the old country, emigrated from eastern Europe in large numbers to America during this period.

Initially, the new immigrants tended to settle where earlier immigrants from their country and members of their family had settled before them. The Lower East Side in Manhattan became a camp of crowded tenements divided along ethnic boundaries: Italian, Jewish, Irish, German. Every great city of the northeastern United States—Philadelphia, Pittsburgh, Boston—had similar concentrations of immigrants joined by ethnicity.

In most cases, the first wave of immigrants retained the language they had brought with them—Jews Yiddish, Italians Italian, Germans German, and so on—at least until their sons and daughters, born and educated in the United States of America, grew up without fluency in the ancestral tongue. The life of the Lower East Side of Manhattan at the turn of the twentieth century was vibrant with newspapers, magazines, social clubs, and theaters in the foreign languages of the immigrants. German-language newspapers, German-medium schools, and German-language churches flourished in areas of heavy German immigration—Wisconsin and central Texas, for example—down to World War I, when war with Germany made pro-German attitudes unpopular and even dangerous. Nonetheless, even into the 1960s one could hear German spoken in America by people who had never set foot in those German-speaking countries from which their grandparents and great-grandparents had emigrated.

America is often referred to as a 'melting pot', implying that different ethnicities merge into one uniform 'American identity' after the first generation of immigrants has assimilated to American life. Most students of ethnicity now have reservations about the melting pot theory. Nathan Glazer and Daniel Patrick Moynihan, in their influential book *Beyond the melting pot* (1963), set the tone for contemporary views when they insisted that 'the point about the melting pot is that it did not happen'. Ethnic minorities have proved to be far more tenacious in clinging to aspects of their Otherness—their ethnic identities—than the melting pot theory assumed. Some minorities, the Amish, for example, or Chasidic Jews, have successfully resisted any but superficial assimilation into mainstream American life.

However, to say that the melting pot is a faulty metaphor for the realities of American assimilation is less true of language than it is of such qualities of ethnicity as religion, cuisine, marriage patterns, and social customs. Foreign languages do not normally survive in America in the second generation except in areas where a large number of immigrants are concentrated, notably Spanish- speakers from many different Spanish-speaking countries in Miami, New York City, southern Texas, and southern California. There are pockets here and there in America where foreign languages defy the rule of second-generation loss, but there are always special circumstances at play in such cases. Examples are the preservation of the Basque language among shepherds in northern California and Québec French in northern Maine. The most common complaint of immigrants who remain in the United States is that their children lose the language; they learn English and rarely have native fluency in the language of their parents.

When the first language of immigration declines, what is often left behind as a residue is an 'ethnic accent'—a pattern of speech influenced by the immigrant language and noticeably different from standard American English. Thus, the pronunciation of *Long Island*, which in standard American pronunciation is [lɔŋ ajlənd] (without a [g]), is often pronounced [lɔŋgajlənd] (with a [g]) in the 'Jewish American accent' of the northeast. This is of course, a stereotype, since by no means do all or most 'Jewish Americans' even in the northeast have this pronunciation; but to the extent that it does occur, it is a carryover from the sound patterns of the Yiddish language. Similarly, *dese* [diz] and *dose* [doz] for *these* [ðiz] and *those* [ðoz] are found in many ethnic accents in which the source language does not have the fricatives [θ] and [ð].

Not all immigrants who came to America were speakers of languages other than English. The Irish began coming to America in large numbers during the period of the Great Famine (1846–1851) when the Irish potato crop failed through disease, and going to America was often the only alternative to death by starvation. Relatively few of these immigrants were

monolingual in the Irish language, the Celtic language indigenous to Ireland (sometimes called 'Gaelic', although 'Irish' is the preferred designation). Most spoke English, but the variety of English they brought with them was 'Irish English', which is different from more 'standard' versions of British or American English.

Characteristics of Irish English pronunciation include: (1) retention of historical /r/ after vowels in words like *car* (as in standard American but not standard British pronunciation); (2) the use of 'clear' /l/ in all positions (i.e. the alveo-palatal [l] of *leaf* vs. the velar [ɫ] of *full*); (3) retention of the contrast between /w/ and /ʍ/ (so that *weather* and *whether*, *witch* and *which* differ in their initial sound); (4) monophthongs /eː oː/ in place of diphthongs [ej ow] in the vowels of words like *face*, *take* and *goat*, *soak*; (5) replacement of the fricatives /θ ð/ by stops and affricates (*thin* [θɪn] pronounced as *tin* [tɪn] or *tthin* [t^θɪn]); (6) retention of vowel distinctions before /r/ (so that the vowels in words such as *bird*, *learn*, *beard*, *turn*, in Irish English, are all different); (7) neutralization of the opposition between /ɛ/ and /ɪ/ before nasals (*pin* and *pen* are homonyms, as in much of southeastern American English). Syntactically, there is the occurrence of reflexive pronouns in sentences such as: *And it's **himself** that told me …* and *… they were paying no attention to anything at all as long as **themselves** were well*. Then, there is the curious matter of the 'after perfect': *I'm after doing it already, She understands; she's after havin' children herself*, and *They seemed pretty cool, for what they were after goin' through* (in place of standard *I've done it, She understands; she's had children herself*, and *They seemed pretty cool after all they had gone through*).

As above in the case of the stereotypical 'Jewish American accent', most Americans of Irish descent have long since accommodated their speech to General American, and none of these Irish-English traits remain in the way they talk. But some of the 'Irishness' would be apt to persist for several generations, especially in areas of big cities such as Boston, New York, and Chicago where the Irish have been resident for a long time and where in many cases the population has been steadily replenished by immigration from Ireland. To the extent that Irish-English speech patterns do persist, they make for what can be called an 'Irish ethnic accent'.

African American Vernacular English (AAVE) has been the most intensively studied ethnic dialect of English. (The terminology is not settled. What is here called AAVE is variously referred to as Black English, African American English, and Ebonics.) AAVE is the dialect of English spoken by African Americans living in the 'inner city'—the ghetto—of large urban centers

such as New York, Philadelphia, Chicago, Detroit, and Los Angeles. The sociolinguist William Labov, in pioneering research undertaken in the 1960s and 1970s, established that AAVE, far from being the 'corrupt, degenerate, ungrammatical, bad' English that linguistically naive observers had thought it to be, is as legitimate a dialect of ethnic English as any other. AAVE is rule-governed and has its own precepts of correctness and incorrectness—it is therefore a language.

Some of the major characteristics of the pronunciation of AAVE are the following: (1) deletion of /r/ except before a vowel (*dough* and *door* rhyme); (2) simplification of final consonant clusters (*he picks* is *he pick*', *she found* is *she foun*'); (3) metathesis (interchange) of consonants in words such as *ask* (*aks*); (4) neutralization of the contrast between [ɛ] and [ɪ] before nasal consonants (compare Irish English above); and (5) replacement of postvocalic [θ] and [ð] by [f] and [v] ([wɪf] for *with*, [brəvɚ] for *brother*). Syntactically, a major systematic difference between AAVE and standard American English is the presence or absence of the verb *be* to distinguish between habitual and momentary state (compare Spanish *ser* and *estar*):

Lisa be sick.	'Lisa is always sick.'
Joe sick.	'Joe is sick at the moment.'
He be late.	'He is chronically late.'
He late.	'He's late this time.'

There is scholarly dispute about the origins of AAVE. The leading opinion at present is that most of the 'nonstandard' features of AAVE can be traced back to the influence of the African languages spoken by the slaves, who were taken from many different tribes and languages of mostly western Africa and acquired English only after they were forced into slavery in America under English-speaking masters and overseers. According to this theory, AAVE arose as a pidgin language based on English in which slaves speaking different tribal languages could communicate. This pidgin language then became a creole language as the children of the slaves grew up speaking it, and from this creole AAVE of the present day has developed.

It must be emphasized that individual differences matter. African Americans, or members of any ethnic group who grow up in contact with standard American English, will speak standard American English. A blindfolded listener could not distinguish their speech from anyone else's. It is a matter of educational opportunities, mobility, and one's language contacts.

Another ethnic dialect of English is Chicano English, the language of Mexican Americans in the southwestern United States and in urban areas to which its speakers have migrated. Characteristics of Chicano

English are: (1) substitution of tense vowels for lax vowels (*sit* and *seat* are both pronounced as *seat*); (2) substitution of *ch* [tʃ] for *sh* [ʃ] (*chew* and *shoe* are both pronounced like *chew* [tʃu]) and vice-versa ([ʃɛk] for *check*); (3) consonant cluster simplification, as in *par'* for *part* and *He like' her* for *He likes her*); and (4) nonstandard patterns of stress and intonation.

A key ingredient of ethnicity is language. Language, that is to say, helps define our place in the world; it can serve either as a sign of membership in the community or as a reason for exclusion from membership in that community. In many parts of the world, ethnic unity and cultural identification are often defined by language rather than by geography or religion. This is notably true of Arabic, whose speakers—the Arabs—base their identity in large measure on the use of a common tongue. The locus of Bengali ethnic identity resides in language despite the division of the speakers of Bengali between two countries, India and Bangladesh, the number of speakers living in Bangladesh being about 100,000,000 and 68,000,000 in India. The Bengali language—a language with an ancient and much-revered literary history—is the principal basis of ethnic unity. This is clear from the name for itself taken by the new nation of eastern Bengal, formerly East Pakistan, following the 1971 war of secession from Pakistan. The noun *Bangladesh* is composed of *bangla* plus *desa*, the latter meaning 'country'. The first part of the compound does not mean the Bengali people or the territory of Bengal; the term *bangla* refers specifically to the Bengali language: *Bangladesh* = 'land of the Bengali speakers'.

The case of the French language in Québec demonstrates a particularly strong association between ethnicity and language. Six million French speakers, five million of them in Québec, compose about one-quarter of the population of Canada. The Québecois (speakers of French resident in the province of Québec) regard themselves as a distinct island in a surrounding sea of Anglophones (monolingual English speakers). They resent the English language and fear its spread in Québec; they resent the historical domination of their economy and culture by Anglophones; and they resent immigrants who want their children to be educated in both English and French. Language is intimately tied to Québecois ethnicity—to Québecois identity. Worship of the French language is the almost sacral force that fortifies and unifies the movement for an independent Québec.

Language is a major symbol of ethnicity, often the major and most tangible symbol. In this role, language has always been a force both for unity and division in the world. It has helped to unify countries—as English makes it possible for Indians from all parts and ethnicities of India to communicate—and it can be a force

for the dissolution of a country. It remains to be seen whether Québec will remain a part of Canada or secede. In Belgium, a virtual language war between French and Flemish (similar to Dutch) threatens this country with its weakly fused identity. Here, as always, ethnicity is bound up with language, economic grievances, and demands for power. The lines of guerilla warfare in Sri Lanka are drawn between Tamil Hindus and Sinhalese Buddhists—*and* between the Tamil and Sinhalese languages. The demands for independence of the Baltic states, Latvia, Lithuania, and Estonia, formerly 'republics' of the Soviet Union, were intimately bound up with fears for the future of their respective ethnicities—and languages—in a sea of Russianness.

One final aspect of the relationship between ethnicity and language deserves attention: the role of language in preserving ethnic identity, especially in diaspora settings. The world has seen countless instances of peoples forced from their homelands because of their ethnicity. In most cases, diaspora leads to loss of language and ultimately of ethnic identity. The celebrated counterexample to this is the role of the Hebrew language in the maintenance of Jewish ethnicity. Hebrew was traditionally the language of the Jewish people; however, it had become extinct as their spoken language by the beginning of the Common Era. It was maintained as the language of ritual, prayer, and disputation among rabbis. After the destruction of the Temple in 70 CE, Jews were dispersed throughout the then known world. Nevertheless, they preserved their identity, Otherness, and distinctive ethnicity. The Hebrew language, which was reborn as a spoken language in Palestine in the nineteenth century, was part of the glue that held Jewish ethnicity together through almost 2,000 years of diaspora.

To paraphrase the great linguist Edward Sapir, we should never make the mistake of confusing a language with a dictionary and a grammar. Both the effect and the affect of language go well beyond words and rules of grammar. Language touches us in the deep places of our being—in our identity, in our sense of where we belong. One of the most sensitive of these places is our ethnicity. In ethnicity begins the true study of language as a badge of identity.

References

Chiswick, Barry R. (ed.) 1992. Immigration, language and ethnicity: Canada and the United States. Washington, DC: AEI Press.

Dow, James R, (ed.) 1991. Language and ethnicity. Amsterdam, Philadelphia: J. Benjamins Publishers.

Fishman, Joshua A. 1989. Language and ethnicity in minority sociolinguistic perspective. Clevedon, Avon, England, Philadelphia: Multilingual Matters, Ltd.

Glazer, Nathan, and Daniel Patrick Moynihan. 1963. Beyond the melting pot. Cambridge, MA: MIT Press.

Gleason, Phillip. 1992. Speaking of diversity: language and ethnicity in twentieth-century America. Baltimore: Johns Hopkins University Press.

Haarman, Harold. 1986. Language in ethnicity: a view of basic ecological relations. Berlin, New York: Mouton de Gruyter.

Lourie, Margaret A., and Nancy Faires Conklin (eds.) 1978. A pluralistic nation: the language issue in the United States. Rowley, MA: Newbury House Publishers.

Shukla, Hira Lal. 1985. Language, ethnicity and history: dimensions in anthropological linguistics. New Delhi, India: B.R. Publishing Corporation.

Van Horne, Winston A. (ed.) 1987. Ethnicity and language. Milwaukee: University of Wisconsin System, Institute on Race and Ethnicity.

ROBERT D. KING

See also **African American Vernacular English; Sociolinguistics**

Ethnography of Communication

Ethnography of communication lies at the intersection of anthropology and linguistics. Until the 1960s, anthropology and linguistics had existed largely in isolation from each other. Although they devoted much of their time to the study of mythology, folklore, and rituals, in which language clearly played an important role, few anthropologists studied language in a systematic manner. In the meantime, few linguists examined the relationship between language and culture systematically. It is against such a background that ethnography of communication came into being. The primary concern of this cross-disciplinary approach is to describe and explain the patterns of communicative behavior of people from a specific culture, across different cultures, or both. The findings of the study are used to illuminate theories of language, as well as theories of culture. Patterns of communicative behavior are regularities that occur so repeatedly that they become predictable. These patterns can be discerned in ritualized events, such as funerals and wedding ceremonies, in which people are often unwittingly compelled to repeat stereotyped expressions. In those not-so-ritualized events, patterns also exist and underlie what people say and how they interact, despite the fact that on the surface participants have every right to choose what to say.

However, patterns are not necessarily universal. Different cultures may have divergent norms governing the same communicative behavior. A typical Chinese way of greeting (e.g. *Qu na?* [literally, 'Where are you going?'] or *Chi le ma?* [literally, 'Have you eaten?']) may sound prying or impolite to an English speaker brought up in a European culture. With a view to recording and analyzing culture-sensitive patterns of communication, proponents of ethnography of communication, such as Dell Hymes, John Joseph Gumperz, Muriel Saville-Troike, and Richard Bauman and Joel Sherzer, have developed an analytic framework by using concepts like 'speech community', 'communicative situation, event, and act', and 'components of speaking'.

A speech community consists of members who not only speak a common language but also interpret the interaction in a similar way. Besides language, members of a speech community share other aspects of common ground, such as social likeness or bonds, nationality, ethnicity, religion, and personal interests. These common grounds can partly account for the fact that most speech communities are bounded geographically, although living in the same region is not a sufficient prerequisite of belonging to the same speech community. The way that speech community members share a common language or norms of interpretation is largely a matter of degree, and it is up to an individual to decide how much commonality is required. A person may simultaneously belong to several discrete or overlapping speech communities. Britain and the United States may, as a whole, be regarded as one speech community if a researcher is interested in the comparison of Western and Eastern cultures. However, when it comes to the comparison of British and American ways of speaking, they have to be treated as two different speech communities rather than as a single unit.

The second essential constituent in the ethnography of communication framework is communicative situation, event, and act. This is a hierarchical concept, with the communicative situation being the general setting (such as a dinner party) and creating the broad context of communication. A communicative situation may consist of one or several communicative events, which refer to specific activities (such as a joke). Communicative events further consist of communicative acts, the minimal unit of analysis (such as greeting).

In ethnography of communication, the key components of analysis are often subsumed under the

mnemonic code SPEAKING, in which each letter represents one aspect of a communicative event. These are the following:

- **S**etting and **S**cene: the concrete physical (e.g. location, time, and size of room) and abstract psychological (e.g. cultural definition of the occasion) circumstances in which communication takes place.
- **P**articipants: the role of each participant in the interaction, their age, sex, social background, and their relationships with one another.
- **E**nds: the purpose and expected outcome of an interaction.
- **A**ct sequence: order of actions, message form, and message content.
- **K**ey: manner and the general tone of interaction.
- **I**nstrumentalities: medium of communication (i.e. spoken or written).
- **N**orms: norms of interaction and interpretation, i.e. what communicative behaviors are regarded as appropriate by a speech community and how communicative behaviors are construed by a speech community.
- **G**enres: categories of communication, e.g. poetry, prayer, or lecture.

Among these components, norms of interaction and interpretation are the most culture specific in that people of a particular culture may share different values or assumptions from other cultures. Consequently, the norms constitute the focus of studies in ethnography of communication. Moreover, the eight components in the list do not hold equal positions in a communicative event. Some elements, such as setting, participants, and key, may determine what norms of interaction and interpretation are to be followed and what act sequence an interaction may take. However, act sequence may tell what norms of interaction and interpretation participants have adopted in conversations, because unlike other variables, norms cannot be directly observed, and their identification has to rely on act sequences, as well as language content, in a conversation.

With its unique analytic framework, ethnography of communication contributes to the understanding of the interrelationship between culture and language by highlighting the role of culture in communicative behaviors. This has important implications for other fields of study, such as psycholinguistics, second-language acquisition, and cross-cultural communication. For psycholinguistics, one of the most significant contributions made by ethnography of communication is that it draws attention to how a child develops a particular way of speaking, apart from innate linguistic knowledge, in the context of a particular culture. For second-language acquisition, the SPEAKING model identifies what a speaker needs to know to communicate appropriately within a particular communicative situation. Learning a second language involves not only acquiring general linguistic knowledge and interaction skills but also social and cultural knowledge that governs what to say to whom and when and how to say it appropriately in relation to communicative goals. For cross-cultural communication, the diversity in the norms of interaction and interpretation in different cultures, emphasized by ethnography of communication, highlights potential sources for misunderstanding and nonunderstanding in cross-cultural interaction. This is also an area that makes ethnography of communication interesting for business consultants and market researchers.

References

Bauman, Richard, and Sherzer Joel. (eds.) 1989. Explorations in the ethnography of speaking. Cambridge: Cambridge University Press.

Gumperz, John Joseph, and Dell Hymes (eds.) 1986. Directions in sociolinguistics: the ethnography of communication. Oxford: Blackwell.

Hymes, Dell. 1962. Foundations in sociolinguistics: an ethnographic approach. Philadelphia: University of Pennsylvania Press.

Saville-Troike, Muriel. 1989. The ethnography of communication: an introduction. Oxford: Blackwell.

ZHU HUA

See also **Psycholinguistics; Sociolinguistics**

Etymology

The field of etymology studies the origin and the earliest meaning of semantic units, i.e. morphemes, words, and phrases. The source (or to use the Greek term, ety-mon) is the word at the time and in the language in which it emerged, i.e. when and where we can break it down into its constituent parts, understand its formation

and the reasons that led to the emergence of its earliest meaning, as well as understand its semantic motivation.

The word *soldier* is borrowed from Old French *soldier*, which in turn developed from Medieval Latin *solidarius* 'one serving in the army for pay'. This is its origin, as in this form the word can be explained as a derivative with the suffix *-arius*, which among other things formed terms for individuals (cf. Medieval Latin *commissarius* > English *commissary* or Medieval Latin *officarius* > (Old) French *officier* > English *officer*), of Medieval Latin *solidus* 'gold coin', which, after applying the respective historical sound change rules is preserved in French *sou* and Italian *soldo* 'wage', pl. *soldi* 'money'. The etymological explanation of the word *soldier* is thus resolved, as the mathematical product of the meanings, i.e. 'paid (military) individual', along with knowledge of the Medieval military situation, corresponds to its formal construction: *solidus* and *-arius*. It is not important for the etymological analysis of the word to indicate that *solidus* in turn is an ellipsis for the collocation *nummus solidus* 'solid coin', nor that the suffix *-arius* was originally adjectival, cf. Latin *manuarius* 'of, or belonging to the hand' from *manus* 'hand'.

The discovery process for etymological explanation of individual words (morphemes or phrases) normally consists of three parts. The first is a synchronic documentation of the word in the language in question and its philologically verifiable development, e.g. *fist* 'the hand closed tightly with the fingers bent against the palm', which is attested with the same meaning in its Old English predecessor *fȳst*. Because the word is attested in the earliest Old English texts, it can be assumed that it existed in the preliterate period, i.e. it is prehistoric. The second step in the etymological process is to establish the direction of further analysis. The question must be asked: did the word arise in a continuous linguistic development—which for English is understood as the development from the prehistoric, preliterate layers Proto-Indo-European > Proto-Germanic > Western Germanic through the historical stages of Old English to Modern English—or was it at some point borrowed from another language, a substratum, adstratum, or superstratum? In view of the fact that the appropriate source form in the languages from which the word might have been borrowed—Romance, Celtic, and Northern Germanic, from which the majority of borrowings came—is lacking, and, on the other hand, the corresponding Western Germanic cognates are attested (cf. Old High German *fūst* [> G. *Faust*], Old Frankish *fest*), it can be posited that the word arose at least in the Proto-West-Germanic period if not earlier. The comparisons at this stage can be affirmed or refuted by the comparative method, which is the fundamental tool of the etymologist. In the example presented here, the Old English phoneme *ȳ*, in light of sound changes established by the comparative method, indicates the existence at an earlier stage of the sound *i* in the second syllable, which was subsequently lost before the historical period. Thus, we can establish that 'fist' was pronounced *fūsti* in the Proto-West-Germanic period.

A word from a prehistoric period, the form and meaning of which are established through the application of the comparative method, is called a reconstruction. An asterisk before the reconstructed form indicates that it is not attested in written documents.

Despite the fact that the reconstructed *fūsti* is not found in other Germanic languages, this form is not yet the origin, as the word cannot yet be analyzed into its original morphemes and thus its original meaning cannot be established. The comparative method allows comparison between the West German reconstruction and Proto-Slavic *pęstь* 'fist', which can be reconstructed from materials in modern and historical Slavic languages, e.g. Slovene *pest*, Russian *pjast'*, Polish *pięść*, and Old Church Slavic *pęstь*. In the comparison, the following comparative-linguistic sound laws are applied: (1) Germanic *f* developed from Indo-European *p* (Grimm's Law); Proto-Slavic preserves Indo-European *p*. (2) Germanic *ū* before *h* developed from *un*, which in turn goes back to Indo-European *n̥*, which in Proto-Slavic normally develops into the nasal vowel *ę*. The development of Germanic long *ū* thus proves that the word is Proto-Germanic and that it goes back to *funhsti-*. The sound *h* in this position could have developed either from Indo-European *k* or *kʷ*, which in the position before another consonant in Proto-Slavic would have disappeared. (3) Proto-Germanic and Proto-Slavic *s* both come from Indo-European *s*. (4) Proto-Germanic and Proto-Slavic *t* both come from Indo-European *t*. (5) The fronting of Old English *ȳ* indicates an older stem ending in *-i*, which is directly affirmed by the Slavic material. The mathematical intersection of all possible reconstructions offered by the Germanic and Slavic material results in the possibilities *pn̥ksti-* and *pn̥kʷsti-*. At this point the comparative-linguistic part of the etymological process ends and the discovery of its origin and analysis into its constituent parts begins. This is done in the framework of Proto-Indo-European word-formational processes.

It turns out that the second of the possibilities mentioned above, *pn̥kʷsti-*, is more likely, as it can be derived from Indo-European *pn̥kʷ-stHi-*, which can be segmented into a compound with the first element *penkʷe* 'five' in the zero-grade and a noun derived from a verb with the root *staH-* 'to stand', also in its zero-grade variant. The etymological process is completed when it is discovered that the original form must have meant 'a standing of five (fingers)'. The explanation is all the more likely as it is typologically

paralleled by *mu-stH-i-* '(small) fist' (attested in Vedic *muṣṭí-* and Tocharian B *maśce*), in which the first element comes from Indo-European *meu̯e* 'four' (Luwian *mauwa-*), and the rest is identical to the Germanic-Slavic word for 'fist'.

Folk etymology is a change in a word or phrase, the constituent parts have become incomprehensible, and as such is reanalyzed with morphemes that are meaningful. The potential for a folk etymology arises when there is a coincidental similarity between the actual morphemes and those—usually with a minor phonological or morphological 'correction'—that speakers see in the word on the principle that 'it makes sense, therefore it must be true'. For example, the common name of a bird *wheatear* (Oenanthe oenanthe) is folk etymologized as a compound of *wheat* and *ear*. This form in turn is a singularized form of an earlier form that had been felt to be a plural, *wheatears*. In fact, the word is a compound from Old English *hwīt* (> Modern English *white*) and *arse*, a reference to the bird's prominent white rump. Folk etymology frequently occurs in borrowed words with minor 'corrections' that seem to speakers to be native derivatives, e.g. German *Abenteuer* 'adventure' is a borrowing of French *aventure*. However, it has been connected with German *Abend* 'evening', apparently on the view that adventures are as a rule evening events. Further, *sparrow grass* is found in the eastern part of the United States as a 'correction' of the unintelligible plant name *asparagus*, itself a borrowing via Latin of Greek *aspáragos*.

References

Malkiel, Yakov. 1976. Etymological dictionaries. A tentative typology. Chicago and London: The University of Chicago Press, ix +144pp.

Malkiel, Yakov. 1993. Etymology. Cambridge: Cambridge University Press, xii + 223pp.

Onions, C. T. 1966. The Oxford dictionary of English etymology. Oxford: Clarendon Press, xvi + 1025pp.

Poruciuc, Adrian. 1991. Folk etymology (in English and elsewhere). Kansas Working Papers in Linguistics 16, pp. 53–8.

MARKO SNOJ

See also **Morpheme; Semantics**

Euphemism

Meanings of words and phrases may be divided into denotation (the referent of the term) and connotation (the associated properties, including social import). If either the denotation or connotation of a word or phrase is offensive to social sensibilities, a 'softer', more polite form may be used. Such forms are *euphemisms*.

When the possible offense in a word lies in its connotation, a word with the same or a similar referent, but lacking this connotation, may be sought. The word 'ofay' refers to Caucasians, but carries with it connotations of opprobrium. If a speaker wishes to denote this group of people, but does not wish to insert negative 'racial' stereotypes, distancing as 'out-group', hatred or dislike, then she or he will use a euphemistic word, such as 'White' or 'Caucasian'. While a given speech community may largely share connotations and can thus count on words 'meaning' the same thing to all in-group interlocutors, connotations are not always shared among all speakers of a language. For some speakers, 'Aryan' might serve as a euphemism for 'ofay'; for others, the connotations of 'Aryan' are also negative, although the non-in-group scope of this term is larger.

While societies differ in their ascription of positive and negative values, all societies have rules of appropriateness of topics for particular groups and settings. In American society, it is generally considered bad form to discuss the body's excretory processes. It has only been in the last 30 years that 'shit' was admitted as an entry in most unabridged dictionaries. Dictionary notes on usage point out that 'shit' is considered 'vulgar'. In polite company, one is expected to refrain from using this word. Should it slip out as an exclamation or as a part of an utterance, one is expected to apologize for the breach of etiquette. In settings, such as medical interviews, where excretory processes must be discussed, euphemisms appear. Biomedical labs refer to 'stool samples'. Nurses inquire about the patient's 'bowel movements'. Euphemisms change with register shifts, as formality levels and statuses of interlocutors vary. Children who must discuss their excretions with the parents during the period of their lives euphemistically known as 'potty' training learn to use terms acceptable to their caregivers, often including 'b.m.', 'doodoo', and 'caca'.

Euphemisms are subject to pejoration. Constant use in reference to something that is itself considered dirty or vulgar pollutes the terms. As long as there is social stigma attached to the referent, some (or all) of it may transfer to the euphemistic term. In the United States, 'nigger' has long been considered a 'hate' word, not apt for polite conversation. In the early 1900s and throughout the civil rights movement in the 1950s, 'colored people' was the phrase of choice. It is preserved today

in the name 'National Association for the Advancement of Colored People', which denotes a social justice organization founded in 1909 by Ida Wells-Barnett, W.E.B. DuBois, Henry Moscowitz, Mary White Ovington, Oswald Garrison Villiard, and William English Walling. In the decade following the 1950s, cultural revitalization groups adopted the term 'black'. It spread from slogans such as 'black is beautiful' to common use as a group label. In the 1970s, Afro-American came to be used. In the next decades, the phrase African American replaced the hyphenated form.

While a society may agree that a topic is 'taboo' or socially loaded, so that reference to it must be oblique, the euphemism of choice may vary, regionally or within social strata. Euphemisms for public places where one may go to 'shit' include: the loo, the w.c., the toilet, the restroom, the washroom, the lavatory, the jensen, the john, the 'ladies', and the 'gents'.

When the discomfort with a term stems from social stresses and when those social schisms become politicized, the lexical choice is politicized as well. Thus, euphemisms are used to achieve 'politically correct' parlance. Feminism has painted the use of generic 'he' and 'man' as indicators of reactionary male hegemonic ideology, so that sentences such as 'Man nurses his young' are sexist, as well as semantically odd.

In those areas where pejoration has proliferated terms, cultural revitalization movements may seek to reclaim elements. In many Latin American countries, the term *indio* 'Indian' has been an insult for centuries; in the last 30 years, some groups in highland Mexico, Guatemala, and Peru are beginning to eschew the euphemistic *indígena* and to self-denominate as *indio*, as a marker of cultural pride. In North American, indigenous groups have been referred to by a variety of euphemistic terms: 'Indian', 'AmerInd', 'Native American', 'First Americans', and 'First Nations'. Amid this multiplicity, one person's euphemism is another's insult or effete affectation.

One can trace areas of conflict, stress, and unease within a society by mapping its euphemisms. Robin Lakoff in her 1975 book *Language and woman's place* notes that female sexuality is one such area of cultural embarrassment. The term 'woman' itself carries sexual connotations; whereas the term 'lady', being more refined, masks these connotations. Assuming that 'Jane' is a 13-year-old girl, compare the following: (a) Jane has become a young lady. (b) Jane has become a young woman. When speakers say that a woman who is a doctor is a 'lady doctor', this indicates not so much her manners, but the fact of her gender, euphemistically stated, and calls attention to it as a 'marked' case, not the cultural norm. As cultural norms shift, the references also shift.

As cultural strife is resolved, euphemisms fall out of play, or lose their sense of masking. As tensions between earlier immigrants and Irish newcomers to the United States subsided, 'Gael' and 'Irish' have fallen out of a euphemistic relation to 'Mick' and 'potato-eater'. As new stresses arise, new lexicon masks and outlines the divisions. Ageism, youth-centrism, and legislation against forced retirement are reflected in the conversion of 'old-age homes' to 'retirement villages', 'nursing homes', and 'assisted living' places for our 'senior citizens' or 'golden agers'. Language forms and is formed by society. Euphemisms are the fresh coat of paint over the stress fractures in the social structure.

References

Allan, Keith, and Kate Burridge. 1991. Euphemism & dysphemism: language used as shield and weapon. New York: Oxford University Press.

Dickins, James. 1998. Extended axiomatic linguistics. Berlin and New York: Mouton de Gruyter.

Enright, D.J. (ed.) 1985. Fair of speech: the uses of euphemism. Oxford and New York: Oxford University Press.

Lakoff, Robin Tolmach. 1975. Language and woman's place. New York: Harper and Row.

Neaman, Judith S., and Carole G. Silver. 1999. The Wordsworth book of euphemism. Lincolnwood, IL: Wordsworth Editions Ltd., NTC Publishing Group.

Spears, Richard A. 1981. Slang and euphemism: a dictionary of oaths, curses, insults, sexual slang and metaphor, racial slurs, drug talk, homosexual lingo, and related matters. Middle Village, NY: David Publishers.

JUDITH M. MAXWELL

European Traditional Grammar

The Greek term *grammatike* (Latin *grammatica*) meant no more at first than the understanding of letters. For the early periods of European history, philosophical works—examining virtually the whole realm of human knowledge—constituted our only sources of grammatical progress.

Plato (c. 429–347 BCE) provided a fundamental division of the Greek sentence into a nominal and a verbal

component, and this remained the primary grammatical distinction underlying sentence analysis and word classification in all subsequent European linguistic descriptions. Aristotle (384–322 BCE) maintained this distinction, but added a third syntactic component, the *syndesmoi*, a class covering what were later to be distinguished as conjunctions, articles, and pronouns.

The Stoic philosophical school, founded by Zeno (c. 315 BCE), developed new methods and doctrines in philosophy and rhetoric and gave separate treatment to grammar, with great achievements: the number of word classes was increased, and more precise definitions and additional grammatical categories were introduced to cover the word and sentence structure associated with these classes. Case in its modern usage as an inflectional category of nouns (and other words inflected like nouns) was the creation of the Stoics. Case inflection became the fundamental division between noun and verb. The Stoics also introduced the distinctions between transitive and intransitive verbs, i.e. verbs that require an object and verbs that do not.

Building on the Stoic ideas, the Greco-Roman world devised technical grammars with an effective and lasting influence on the future. The framework of technical grammar in antiquity was the *word-and-paradigm* model. A word-based grammar involves three main procedures: the identification of the word as an isolable linguistic entity, the establishment of a set of word classes to distinguish and classify the words in the language, and the working out of adequate grammatical categories to describe and analyze the structure and formation of words (morphology) entering into paradigms of associated forms.

The first piece of technical grammar in Europe was the *Tekhne Grammatike* of Dionysius Thrax (c. 100 BCE). Its first part is considered genuine; the rest is probably a later addition. The grammarian classified the consonants of Greek and distinguished eight word classes: noun, verb, participle, article, pronoun, preposition, adverb, and conjunction. The number of word classes, with one change necessitated by the absence of articles in Latin, remained constant until the end of the Middle Ages in the grammatical description of Greek and Latin, and it had a very marked influence on the grammatical analysis of several modern European languages. Each word class defined in the *Tekhne Grammatike* is followed by a statement of the categories applicable to it: gender (masculine, feminine, or neuter), type (primary or derived), form (simple or compound), number (singular, dual, or plural), and case (nominative, vocative, accusative, genitive, or dative). The classification categories introduced for the verb are mood, voice, type, form, number, person, tense, and conjugation.

The main omission of the *Tekhne* from the standpoint of modern linguistics was the absence of any sections on syntax, i.e. sentence structure. Syntax was dealt with extensively by Apollonius Dyscolus (second century CE). He sharply distinguished form and meaning, and he justified grammatical classifications by reference to meaning rather than to form. His work, however, is a case study of Greek, rather than a general theory of syntax.

Varro (first century BCE) is the first important Latin author with linguistic concerns whose work has been transmitted. In the surviving part of his *De Lingua Latina*, he distinguishes word formation processes in which the class of the word is changed (e.g. derivation of a noun from a verb via the addition of an affix, such as the noun 'runner' from the verb 'to run') and class-preserving inflections (e.g. tense markings on the verb, such as 'moved' for the past tense of 'move'). This distinction between derivational and inflectional processes was not commonly made in antiquity. Varro also proposed his own classification of Latin-inflected words by setting up a system of four contrasting classes of inflections.

Aelius Donatus was the most influential Latin grammarian of the fourth century CE and dominated grammatical doctrine in Europe at least until the twelfth century. His *Ars minor* deals with the eight parts of speech in a question-and-answer format. His *Ars maior* is more comprehensive and includes style guidelines for Latin.

Priscian's voluminous grammar (c. 500 CE) was the most important late Latin grammar. He was under the spell of Greek grammar, accepting Apollonius and his son Herodian as authorities and using the closest Latin translations for the Greek technical terms. Priscian organized the morphological description of the forms of nouns and verbs and of the other inflected words by setting up canonical, or basic, forms; from these, he derived the other forms by a series of letter changes, the letter being for him, as for the rest of Western antiquity, both the minimal graphical unit and the minimal phonological, or sound, unit.

In the first centuries after the collapse of the Western Roman Empire, grammar served mainly practical and normative purposes. Later writings took the form of commentaries and glosses, mainly based on Priscian. Donatus and Priscian shared the place of the principal grammarians of the Middle Ages as teachers' textbooks. In England, Bede and Alcuin wrote grammars of Latin in the seventh and eighth centuries, following Donatus.

Until the twelfth century, Greek and Latin were regarded as the pinnacle of culture, which explains the grammarian's exclusive attention to these languages. In the latter part of the Middle Ages, however, descriptions of other European languages appeared, serving the ends of literacy, popular literature, and educational standards. Irish grammarians showed a great deal of originality in their treatment of Irish; descriptions of

Provençal started appearing after 1240; an anonymous grammarian of the period demonstrated great insight into the description of Icelandic, especially in phonology and its graphemic representation.

The Renaissance saw the first European grammars of Hebrew and Arabic. Interest in these languages and the separate scholarly traditions in which they had been treated contributed to further loosening the strict focus on Latin and Greek. The first known native grammars of Spanish and Italian appeared in the fifteenth century, followed closely by French, Polish, and Old Church Slavonic grammars. The first printed grammar of English appeared in 1586.

The new grammars of modern languages paid great attention to the relations between spelling—now being standardized in printing—and pronunciation. Although the confusing equation of letter and spoken sound continued, phonological inadequacies of existing spelling were noted and criticized. The system of eight word classes was challenged, with systems of fewer and more classes being proposed.

English grammarians of the sixteenth, seventeenth, and eighteenth centuries usually started with the Latin system of the eight Priscianic classes. They either followed it or felt the need to express and justify their disagreements with it. The English articles, *a, an,* and *the*, having no Latin counterpart, were not given the status of a part of speech, but were merely referred to as notes or signs set before nouns to identify them as nouns. Others treated the articles as adjectives (still a subcategory of nouns), and Ben Johnson (1640) assigned them to a class of their own. The influence of the Latin tradition is seen in the retention of the adjective within the noun class, although there are fewer reasons for this theoretical choice in English than in Latin. The preoccupation of most of the grammarians with the participle, treated either (purely traditionally) as a class in its own right or as a noun adjective having particular derivational associations with the verb, is also a carryover from Latin grammar.

In the New World, the first grammar of a Native American language, Tarascan (Mexico), was published in 1558; thereafter, grammars of Quechua (Peru), Nahuatl (Mexico), and Guarani (Brazil) were published in 1560, 1571, and 1640, respectively. The seventeenth century witnessed the publication of grammars of Japanese and Persian, and the first comprehensive grammars of Chinese published in European languages appeared at the beginning of the eighteenth century. Confidence in the traditional Greco-Roman grammatical categories was weakened when it became known that Chinese generally uses independent words where Greek and Latin use affixes, whereas Native American languages often use the opposite strategy of expressing the equivalent of Greek or Latin sentences in highly complex words.

The 'discovery' of Sanskrit and the emergence of the so-called Neogrammarian theories in the nineteenth century, however, represented the most serious challenges, both to the concepts of European traditional grammar and to the idolization of Greek and Latin. The fact that Sanskrit was found to be closely related to Greek and Latin stripped them of their attributed uniqueness, and the Neogrammarian focus on the development of languages over time put them into historical perspective.

References

Keil, Heinrich (ed.) 1857–1870. Grammatici Latini, 8 vols. Leipzig: Teubner.

Kemp, Alan. 1987. The Tekhne Grammatike of Dionysius Thrax: English translation with introduction and notes. The history of linguistics in the Classical Period, ed. by Daniel J. Taylor. Amsterdam and Philadelphia: John Benjamins.

Law, Vivien, and Ineke Sluiter (eds.). Dionysius Thrax and the Techne Grammatike. The Henry Sweet Society studies in the history of linguistics 1. Münster: Nodus Publikationen.

Lepschy, Giulio (ed.) 1994, 1998. Storia della linguistica, Bologna: Il Mulino; as History of linguistics: Vol. II: classical and medieval linguistics, Vol. III: renaissance and early modern linguistics. London and New York: Longman.

Robins, R.H. 1967. A short history of linguistics. London and New York: Longman, 4th edition, 1997.

Taylor, Daniel J. (ed.) 1987. The history of linguistics in the classical period. Amsterdam and Philadelphia: John Benjamins.

Schneider, R., G. Uhlig, and A. Hilgard (eds.) 1867–1910. Grammatici Graeci. Leipzig: Teubner.

CHRISTOS G. NIFADOPOULOS

See also **Grammar, Traditional; Grammar, Theories; History of Linguistics: Overview; Morphology**

Evolution of Language 1: Overview

Although the details are disputed, it is apparent that human linguistic ability has a biological basis. That being the case, any meaningful study of the evolution of language must take into account the principles noted by Charles Darwin; they remain central tenets of evolutionary biology.

The evolutionary mechanism most often associated with Darwin is natural selection, which acts on 'any variation, however slight and from whatever cause proceeding, if it be in any degree profitable to an individual of any species, in its infinitely complex relations to other organic beings and to external nature…' (1859; *On the origin of species,* p. 61). Because human language enhances virtually all aspects of our interactions with each other and external nature, selection for the biological characteristics that make language more effective is to be expected.

Comparative studies of primates suggest that language may foster group cohesion and the formation of large social groups working cooperatively by means of seemingly useless 'gossip'. Other studies note that language allows information to be shared by many individuals, thereby fostering the solution of problems through 'collective insight'. Robust experimental data show that human speech itself is one of the keys to human linguistic ability. The process by which human speech is produced and perceived overcomes the slow information transfer rate of the human auditory system. As we talk, phonemes (speech sounds roughly equivalent to the letters of the alphabet) are melded into syllables that are transmitted below the rate at which the human auditory system would have fused individual sounds into an incomprehensible buzz. Studies of sentence comprehension reveal another role for speech. Words are silently 'rehearsed' by using the neural mechanisms that regulate overt speech, to maintain them in the neural buffer, *verbal working memory,* in which sentence comprehension takes place. Other avenues of inquiry that use computer modeling techniques suggest that the complex syntax of human language derives from limited neural memory and sentence-processing capacities.

Darwin also introduced the comparative method, which can yield reasonable inferences concerning the evolution of the 'derived' characteristics that differentiate a living species from its extinct ancestors. For example, comparative studies of living apes and humans show that upright bipedal locomotion involves certain derived anatomical features that can be traced back in time by examining the fossil record. In this light, studies of chimpanzees shed light on the evolution of human language. Although present-day chimpanzees are not members of the extinct 6- to 7-million-year-old species that was the common ancestor of humans and apes, they preserve many of the skeletal features of early hominids, including small brains and similar vocal anatomy. However, chimpanzees raised in human-like environments can acquire approximately 150 words and the ability to comprehend sentences that have a simple syntax with manual languages (American Sign Language or computer keyboards). This suggests that limited lexical and syntactic ability were present in the earliest stages of the evolution of human language and argues against theories that posit an abrupt transition that yielded human syntactic ability. Natural selection could have gradually enhanced preexisting rudimentary lexical and syntactic ability. In contrast, the fact that chimpanzees totally lack voluntary speech suggests that the speech capabilities of early hominids were limited, and this is consistent with studies of the skulls of extinct hominids. These skulls suggest a chimpanzee-like vocal anatomy incapable of producing the full range of human speech sounds.

The archeological record also preserves artifacts that hint at the cognitive and linguistic abilities of ancestral and related extinct hominid species. However, we must take into account the effects of the accelerating pace of human culture. Although we can fly through the air, our cognitive and linguistic ability is not superior to that of our ancestors during the tens of thousands of years in which horses provided the fastest means of land transportation. Moreover, genetic data must temper conclusions based on the archeological record. Although tool technologies seem to have abruptly changed 40,000 years ago in Europe, it is improbable that modern human linguistic ability first evolved at that date. Genetic studies and the fossil record show that modern human beings left Africa about 100,000 years ago and settled in most parts of the world between 60,000 and 40,000 years ago. Because any child from any region of the world can acquire any language with native proficiency, the biological capacities for language must have existed before humans left Africa.

But some linguists question how human language could have evolved by means of natural selection in the brief 5-million-year period separating humans and living apes. As Darwin himself noted, natural selection can 'act only by the preservation and accumulation of infinitely small inherited modifications' (1859:95). How then could he account for major adaptations to new environments, such as the transition from aquatic to terrestrial life? Darwin's solution, often termed 'exaptation' or 'preadaptation', rested on the observation that 'an organ might be modified for some other and quite distinct purpose' (1859:190). Converging evidence suggests that preadaptation played a part in the evolution of the unique characteristics of human language, such as voluntary speech and complex syntax. The anatomy of the face and tongue was modified to allow the production of sounds that make human speech more efficient. Brain structures initially adapted for motor control may have been the starting point for neural networks that regulate syntax and cognition.

In studying the evolution of linguistic ability, general evolutionary principles, such as Von Baer's 'law',

provide another avenue of inquiry. Von Baer's law holds that the ontogenetic development, from fetal life onward, of the derived anatomical and related behavioral characteristics of a living species generally follows a sequence that mirrors their occurrence in related living and extinct species. Inferences on the evolution of the anatomy of speech production, syllable structure, vocabulary ability, and sentence structure have followed from studies of the development of children.

The exact course of the evolution of human language will probably never be known, because detailed timetables cannot be subjected to test. However, because the mark of evolution is apparent when we study the human brain and body, further insights into the evolution of human language will become apparent as our knowledge of the biological basis of human language and cognition advances.

References

Darwin, Charles. 1859. On the origin of species, London: Murray, Cambridge, MA: Harvard University Press, Facsimile edition, 1964.

Deacon, Terrance W. 1997. The symbolic species: the co-evolution of language and the brain. New York: W. W. Norton.

Donald, Merlin. 1991. Origins of the modern mind. Cambridge, MA: Harvard University Press.

Gardner, R. Allan, Beatrix T. Gardner, and Thomas A. Van Cantfort. 1989. Teaching sign language to chimpanzees. Albany: State University of New York Press.

Gibson, Katherine R., and Thomas Ingold. 1993. Tools and language in human evolution. Tools, language and cognition in human evolution. Cambridge: Cambridge University Press.

Hauser, Marc D. 1996. The evolution of communication. Cambridge, MA: MIT Press.

Kirby, Simon, and Morton Christiansen (eds.) 2001. Language evolution. The states of the art. Oxford: Oxford University Press.

Lieberman, Philip. 1984. The biology and evolution of language. Cambridge, MA: Harvard University Press.

Lieberman, Philip. 1998. Eve spoke; human language and human evolution. New York: W. W. Norton, London: Picador, Macmillan.

Lieberman, Philip. 2000. Human language and our reptilian brain: the subcortical bases of speech, syntax, and thought. Cambridge, MA: Harvard University Press.

Mellars, Paul. 1996. The Neanderthal legacy: an archaeological perspective from Western Europe. Princeton, NJ: Princeton University Press.

Mellars, Paul, and Christopher B. Stringer. 1990. The human revolution: behavioral and biological perspectives on the origin of modern humans, Vol. 1. Edinburgh: Edinburgh University Press.

Savage-Rumbaugh, Sue, and Roger Lewin. 1994. Kanzi: the ape at the brink of the human mind. New York: Wiley.

PHILIP LIEBERMAN

See also **Animals and Human Language 4: Primates; Evolution of Language 2: Cognitive Preadaptations; Evolution of Language 3: Physical Preadaptation; Evolution of Language 4: Social Preadaptations**

Evolution of Language 2: Cognitive Preadaptations

A preadaptation is a change in a species which is not itself adaptive but paves the way for subsequent adaptive changes. For example, bipedalism set in train anatomical changes, which culminated in the human vocal tract. Although speech is clearly adaptive, bipedalism is not itself an adaptation for speech; it is a preadaptation. This example involves the hardware of language, the vocal tract. Many changes in our species' software, our mental capacities, were necessary before we became language-ready; these are cognitive preadaptations for language.

A human language is a system mapping combinations of speech sounds or manual signs onto meanings. The use of language is a cooperative, social activity. When humans use language, they cause inferences in a receiver's mind about states of affairs in their shared world. Preadaptations for language involved the following capacities or dispositions:

pre-phonetic capacity to produce speech sounds or manual gestures.

pre-syntactic capacity to organize longer sequences of sounds or gestures.

pre-semantic capacities to: (a) form basic concepts, (b) construct more complex concepts (e.g. propositions), and (c) carry out mental calculations over complex concepts.

pre-pragmatic capacities to: (a) infer what mental calculations others can carry out, (b) act cooperatively, (c) attend to the same external situations as others, and (d) accept symbolic action as a surrogate for real action.

311

elementary symbolic capacity to link sounds or gestures arbitrarily with basic concepts, such that perception of the action activates the concept, and attention to the concept may initiate the sound or gesture.

If some capacity is found in species distantly related to humans, this indicates that it is an ancient, primitive capacity. Conversely, if only our nearest relatives, the apes, possess some capacity, we can conclude that it is a more recent evolutionary development. Twin recurring themes in the discussion of many of these abilities are *learned*, as opposed to innate behavior, and *voluntary control* of behavior. Human languages are largely learned systems. The more ways in which a species is plastic in its behavior, the more complex are the cultural traditions, including languages, that can emerge. Our nearest relatives, the chimpanzees, are plastic in a significantly wider range of behaviors than any other nonhuman animals; their cultural traditions are correspondingly more multifaceted while falling far short of human cultural diversity and complexity. Combined with plasticity, voluntary control adds more complexity, and unpredictability, to patterns of behavior. Much of the difference between humans and other species can be attributed to greatly increased plasticity and voluntary control of these preadaptive capacities.

Prephonetic Capacity

Chimpanzees cannot speak. They typically have little voluntary breath control. To wild chimpanzees, voluntary breath control does not come naturally. On the other hand, chimpanzees have good voluntary control over their manual gestures, although they are not as capable as humans at delicate manual work. A preadaptation that was necessary for the emergence of modern spoken language was the extension of voluntary control from the hands to the vocal tract.

Learning the controlled use of physical actions entails an ability to imitate. Imitation involves an impressive 'translation' of sensory impressions into motor commands. Think of a smile. Without mirrors or language, one can have no guarantee that a certain set of muscle contractions produces the effect one perceives in another's face. Given the required voluntary control, and the anatomical hardware, imitation of speech sounds should be easier than imitation of facial gestures, because one can hear one's own voice. A capacity for imitation is found in a perplexing range of species. Some birds can imitate human speech, and many other types of sound, as well. Dolphins can be trained to imitate human movements. A capacity for imitation can evolve separately in different species, with or without the other necessary preadaptive requirements for human language. A neural basis of

imitation has been found in monkeys, in the form of 'mirror neurons', which fire both when an animal is carrying out a certain action, such as grasping, and when it observes that same action carried out by another animal. A recurrent theory in phonetics is the 'motor theory of speech perception', which claims that speech sounds are represented in the brain in terms of the motor commands required to make them.

Although they cannot speak, our ape cousins have no trouble in recognizing different spoken human words. The capacity to discriminate the kinds of sounds that constitute speech evidently preceded the arrival of speech itself.

Presyntactic Capacity

Syntax is the stringing together of independent subunits into a longer signal. The oscine birds, in which learned complex song is observed, are extremely distant relatives of humans. Many other birds, and more closely related species, including most mammals, do not, as far as we know, produce calls composed of independent subunits. Our closest relatives, the apes, do produce long calls composed of subunits. The long calls of gibbons are markers of individual identity, for advertising or defending territory. The subunit notes, used in isolation, out of the context of long calls, are used in connection with territorial aggression, and it is not clear whether the meanings of these notes can be composed by any plausible operation to yield the identity-denoting meaning of the whole signal.

Male gibbon singing performances are notable for their extreme versatility. Precise copies of songs are rarely repeated consecutively, and the song repertoires of individual males are very large. Despite this variability, rules govern the internal structure of songs. Male gibbons use a discrete number of notes to construct songs. Songs are not formed through a random assortment of notes. The use of note types varies as a function of position, and transitions between note types are nonrandom. (Mitani and Marler 1989:35)

Although it is fair to call such abilities in apes 'presyntactic', they are still far removed from the human ability to organize sequences of words into complex hierarchically organized sentences. Little is yet known about the ability of apes to learn hierarchically structured behaviors, although all researchers seem to expect apes to be less proficient at it than humans.

Presemantic Capacities

Basic Concept Formation

Many distantly related species lead simple lives, compared to humans, and even to apes, and so may not possess very many concepts, but they do nevertheless

possess them. 'Perceptual categorisation and the retention of inner descriptions of objects are intrinsic characteristics of brain function in many other animals apart from the anthropoid apes.' (Walker 1983:378). The difference between humans and other animals in terms of their inventories of concepts is quantitative. Animals have the concepts that they need, adapted to their own physiology and ecological niche. What is so surprising about humans is how many concepts they have, or are capable of acquiring, and that these concepts can go well beyond the range of what is immediately useful. Basic concrete concepts, constituting an elementary presemantic capacity, were possessed by our remote ancestors.

Something related to voluntary control is also relevant to presemantic abilities. We do not need to be stimulated by the actual presence of an object for a concept of it to be brought to mind. Some animals may have this to a limited degree. When an animal sets off on a journey to its usual foraging ground, it knows where it is going, because it can get there from many different places, and even take new routes. Hence, the animal is entertaining a concept of a place other than where it currently is. But for full human language to have taken off, a way had to evolve of mentally reviewing one's thoughts in a much more free-ranging way than animals seem to use.

Complex Concept Formation
The ability to form complex conceptual structures, composed systematically of parts, is crucial to human language. Logical predicate-argument structure underlies the messages transmitted by language. The words comprising human sentences typically correspond to elements of a conceptual/logical representation. While apes may perhaps not be capable of storing such complex structures in their heads as humans, it seems certain that they have mental representations of the predicate-argument form. Simply attending to an object is analogous to assigning a mental variable to it, which functions as the argument of any predicate expressing a judgment made by the animal. The two processes of attending to an object and forming some judgment about it are neurologically separate, involving different pathways in the brain. This is true not only for humans but also for apes and closely related monkeys as well. It seems certain that all species closely related to humans, and many species more distantly related, have at least this representational capacity. This capacity is a presemantic preadaptation for language.

Mental Calculation
Humans are not the only species capable of reasoning from experienced facts to predictions about nonexperienced states of affairs. There is a large literature on problem-solving by animals, leading to ranking of various species according to how well they perform in some task involving simple inference from recent experience. Apes and monkeys perform closest to humans in problem-solving, but their inferential ability falls short of human attainment.

Prepragmatic Capacities

Mind-reading and Manipulation
When a human hears an utterance by another person, he often has to figure out what the speaker intended; this is mind-reading. When a human speaks, he or she usually does so with some estimation of how the hearer will react; this is social manipulation. Humans have especially well-developed capacities for social manipulation and mind-reading, and these evolved from similar abilities in our ancestors, still visible in apes. Social intelligence, a well-developed ability to understand and predict the actions of fellow members of one's group, was a necessary prerequisite for the emergence of language. Recent studies amply demonstrate these manipulations and mind-reading abilities in chimpanzees.

Cooperation
People are able to understand the intended import of statements whose literal meanings are somehow inappropriate, such as 'It's cold in here', intended as a request to close the window. To explain how we cope with such indirectness, traditional logic has to be supplemented by the Cooperative Principle, which stipulates that language users always try to be helpful in various specified ways. The use of language requires this basis of cooperativeness. No such complex communication system could have evolved in the absence of assumed cooperativeness between members of the community.

Humans are near the top of the range of cooperativeness, compared with other species. The basis of cooperation in social insects is entirely innate, and the range of individual cooperative behaviors is small. In humans, building onto a general natural disposition to be cooperative, cooperation on a wide range of specific group enterprises is culturally transmitted. Children are taught to be 'team players'. No concerted instruction in cooperation exists outside humans, but there are reports of cases where an animal appears to be punished for some transgression of cooperativeness. So the basis for cooperative behavior, and for instilling such behavior in others, exists in species closely related to humans. Common Chimpanzees and bonobos, in particular, frequently engage in many different types of reconciliation and peace-making behavior. Dispositions to cooperation and maintenance of group cohesion are pragmatic cognitive preadaptations for language.

Joint Attention

Cats are hopelessly inept at following a pointing finger; dogs are somewhat better. Language is also used to point at things, both directly and indirectly. Linguists and philosophers call this 'reference'. When a speaker refers to some other person, say by using a personal pronoun, such as *he*, the intention is to get the hearer to attend to this other person. Successful use of language demands an ability to know what the other person is talking about. A mechanism for establishing joint attention is necessary. Human babies and children are especially adept at gaze- and finger-following. The fact that humans, uniquely, have whites in their eyes, probably helps us to figure out what other people are looking at.

Primates more closely related to humans are better at following the human gaze than those less closely related. Chimpanzees follow human gaze cues, while non-ape species such as macaques fail to follow human gaze cues. But experiments on rhesus macaques interacting with other rhesus macaques show that members of this species do follow the gaze of other members of their own species. Spontaneous pointing has also been observed in captive common chimpanzees (who had not received language training) and in young free-ranging orangutans. It thus appears that animals close to humans possess much of the cognitive apparatus for establishing joint attention, which is the basis of reference in language.

Ritualized Action

Short colloquial greetings such as *Hello!* and *Hi!* are just act-performing words; they do not describe anything, and they can not be said to be true or false, as a description can. We can find exactly such act-performing signals in certain ritualized actions of animals. The classic example of a ritualized action is the snarling baring of the teeth by dogs, which need not (now) precede an imminent attack, and is a sign of hostility. In order to bite, a dog must get its lips out of the way. Originally, parting the lips was only a preparatory movement before biting, but observers became aware of the baring of teeth as an index of an attack and took evasive action. Teeth-barers soon noticed that simply baring the teeth brought about the desired effect of scaring someone off. Now, baring the teeth is a ritualized signal with a meaning something like 'Beware!'. Human ritualized expressions such as *Hello* are relics of ancient animal behavior, now mostly clothed in the phonemes of the relevant language. But some human ritualized expressions, such as the alveolar click, 'tsk', indicating disapproval, are not assimilated into the phonology of their language (in this case English). The process of dissociation between the form of the signal and its meaning can be seen as the basis of the capacity to form arbitrary associations between signals and their meanings, to be discussed in the next section.

Elementary Symbolic Capacity

The sound of the word *tree*, for instance, has no iconic similarity to any property of a tree. This kind of arbitrary association is central to language. Human linguistic symbols are entirely learned. This excludes from language proper any possible universally instinctive cries, such as screams of pain or whimpers of fear. In the wild, there are many animals with limited repertoires of calls indicating the affective state of the animal. In some cases, such calls also relate systematically to constant aspects of the environment. The best-known example is the vervet monkey alarm system, in which there are distinctive calls for different classes of predator. There is no evidence that such calls are ever learned to any significant degree. Thus, no animal calls, as made in the wild, can, as yet, be taken as showing an ability to learn an arbitrary mapping from signal to message.

Trained animals, on the other hand, especially apes, have been shown to be clearly capable of acquiring arbitrary mappings between concepts and signals. The acquired vocabularies of trained apes are comparable to those of five-year-old children, with hundreds of learned items. An ape can make a mental link between an abstract symbol and some object or action, but the circumstances of wild life never nurture this ability, and it remains undeveloped.

The earliest use of arbitrary symbols in our species was perhaps to indicate personal identity. They replaced nonsymbolic indicators of status such as physical size, and involuntary indexes such as plumage displays. In gibbons, territorial calls also have features that can indicate sex, rank, and (un)mated condition.

The duetting long call behavior seen in chimpanzees and bonobos, in which one animal matches its call to that of another, indicates a degree of transferrability of the calls between individuals, and an element of learning. But it seems likely that such duetting is 'parrot-like', in that the imitating animal is in no way attempting to convey the 'meaning' (e.g. rank, identity) of the imitated call. The duetting behavior is not evidence of transfer of *symbolic* behavior from one individual to another. Probably, the duetting behavior itself has some social/pragmatic significance, perhaps similar to grooming.

In humans the ability to trade conversationally in symbols comes naturally. Even humans have some difficulty when the symbol clashes with its meaning, as, for example, if the word 'red' is printed in green. Humans are able to overcome such difficulties and get

a response to the *symbol* to take precedence over the response to the *thing*. But chimpanzees apparently cannot suppress an instinctive response to concrete stimuli in favor of response to symbols. With few exceptions, even trained apes usually only indulge in symbolic behavior to satisfy immediate desires. The circumstances of wild chimpanzee life have not led to the evolution of a species of animal with a high readiness or willingness (as with humans) to use symbols, even though the rudiments of symbolic ability are present.

References

Byrne, Richard W., and Andrew Whiten. 1988. Machiavellian intelligence: social expertise and the evolution of intellect in monkeys, apes and humans. Oxford: Clarendon Press.

de Waal, Frans B.M. 1989. Peacemaking among primates. Cambridge, MA: Harvard University Press.

Donald, Merlin. 1991. Origins of the modern mind: three stages in the evolution of culture and cognition. Cambridge, MA: Harvard University Press.

Hauser, Marc D. 1996. The evolution of communication. Cambridge, MA: MIT Press.

Hurford, James R., Michael Studdert-Kennedy, and Chris Knight (eds.) 1998. Approaches to the evolution of language: social and cognitive bases. Cambridge: Cambridge University Press.

Jolly, Alison. 1985. The evolution of primate behavior, 2nd edition. New York: Macmillan.

Knight, Chris, Michael Studdert-Kennedy, and James R. Hurford (eds.) 2000. The evolutionary emergence of language: social function and the origins of linguistic form. Cambridge: Cambridge University Press.

Mitani, J.C., and Peter Marler. 1989. A phonological analysis of male gibbon singing behavior. Behaviour 109.

Walker, Stephen. 1983. Animal thought. London: Routledge and Kegan Paul.

JAMES R. HURFORD

See also **Animals and Human Language 3: Parrots; Animals and Human Language 4: Primates; Evolution of Language 1: Overview; Evolution of Language 3: Physical Preadaptation; Evolution of Language 4: Social Preadaptations**

Evolution of Language 3: Physical Preadaptations

We owe the concept of *preadaptation* (sometimes termed 'exaptation') to Charles Darwin (1859; *On the origin of species*: 190), who observed that 'an organ might be modified for some other and quite distinct purpose'. Preadaptation explains many seemingly unaccountable transitions in the evolutionary process. For example, the bones of the mammalian middle ear derive from the hinged jawbones of reptiles, and the retina of the eye derives from normal skin tissue. Preadaptation is responsible for many of the characteristics that differentiate human linguistic ability from the communications of closely related living species, such as chimpanzees.

In this light, studies of the communicative abilities of chimpanzees are of special interest. Approximately 98% of the genetic code is common to humans and chimpanzees; the skeletal remains of early, extinct hominids (species close to or in the line of human evolution) are similar in many ways to those of chimpanzees. Therefore, any aspect of language shared by humans and chimpanzees is most likely a 'primitive' characteristic that characterized the linguistic ability of our common ancestors some 6 to 7 million years ago. And features that differentiate humans from chimpanzees must have evolved during that period. Studies of the communications of chimpanzees in situations where they have been raised by humans and exposed to manually transmitted languages (American Sign Language or computer keyboards that code words) show that they possess the ability to acquire about 150 words and can comprehend spoken English sentences that have a simple syntax. In contrast, chimpanzees cannot talk, nor can they comprehend or produce (by using sign language) utterances that have complex syntax. We can conclude that speech production is a derived feature of human language.

Preadaptation is involved in the evolution of virtually all aspects of speech production. Speech production involves three components: (1) the lungs, which generate the expiratory airflow that is the power source; (2) the larynx, which serves for phonation, i.e. it converts alveolar airflow through vibration into a quasiperiodic series of puffs of air containing acoustic energy at audible frequencies; and (3) the airway between the larynx and lips, the supralaryngeal vocal tract (SVT), which acts as an acoustic filter, allowing maximum acoustic energy to pass through it at certain *formant* frequencies.

Speech has an overt linguistic function closely linked to syntax. The flow of air from the lungs can segment the flow of speech into sentences or into shorter clause-length segments. Human speakers can mark out the units for syntactic analysis, usually sentences, by producing each one with a single breath. A constellation of acoustic cues marks the end of each breath group. The fundamental frequency of phonation (F0), the rate at which the larynx produces puffs of air (perceived as the *pitch* of a speaker's voice), and distinctions in syllable duration and amplitude, signal the end of a breath group. These acoustic cues are the primary physical markers of sentence *intonation*. Long sentences can be linked by using two or more breath groups. The fact that speech is produced and segmented by expiratory airflow follows from the preadaptive, evolutionary history of our lungs. As Charles Darwin noted in 1859, the lungs of terrestrial animals derive from the swim bladders of fish. The elasticity of the lungs, which reflects their original function, lends itself to the production of expiratory airflow patterns that have a relatively steady air pressure. During spontaneous speech, human speakers generally anticipate the sentence that they intend to speak and execute a complex set of motor gestures to produce a breath group that has the appropriate length. The neural capacity to execute breath groups and intonation keyed to sentences may be a unique human *derived* capacity. Preliminary studies of the nerve channels of one specimen of *Homo erectus* suggest that this capacity did not exist 1 million years ago—perhaps not until the epoch of Neanderthals and modern humans.

The larynx's linguistic functions likewise show the mark of preadaptation. The original function of the larynx was to seal off the lungs of lungfish when they were underwater. In frogs and many species ultimately related to frogs (including humans), the larynx has been modified for phonation. Variations in the fundamental frequency (F0) convey emotion and affect in humans and in many other species. In many languages, systematic F0 variations alone serve to differentiate words, for example, [ma] produced with the four different *tones* of Mandarin Chinese yields four different words.

The human SVT generates the formant frequencies that differentiate vowels and consonants. The human SVT also shows the mark of preadaptation. In all primates and most mammals, including human beings older than three months, the airway above the larynx is adapted for simultaneously ingesting liquids and swallowing small chunks of solid food during breathing. In newborn primates, including humans, a relatively thin tongue is positioned almost entirely within the mouth; the larynx is positioned close to the opening to the nasal cavity and can be raised (periscope fashion), pro-

viding a sealed nose-to-lung air path during drinking. The distance between the back of the hard palate (the roof of the mouth) and vertebral column is relatively long to accommodate the raised larynx. During the first year of human life, the face moves backward relative to the rest of the skull, and the larynx descends. The larynx continues to descend until age five to six years, when the length of the mouth (the oral cavity) and the pharynx (the vertical section of the SVT) are almost equal. The pharynx and oral cavities are almost at right angles. During the course of this process, the posterior contour of the human tongue becomes almost round, forming the anterior (front) wall of the pharynx. During the production of speech, the human tongue is moved about, almost undeformed, in the right-angle space defined by the oral cavity and pharynx.

In contrast, apes follow a different growth process. As they mature, their faces restructure and project forward. They retain the high laryngeal position, long mouths, and SVTs of newborn primates. The skulls of early extinct hominids, such as Australopithecines, seem to have retained the ape-like SVTs of the common ape and human ancestral species. Examination of the skulls of juvenile and adult Neanderthal hominids, who diverged from modern human beings 500,000 years ago, show that they followed a modified nonhuman growth trajectory.

The human SVT maintains this approximately 1:1 oral/pharynx proportions throughout life. The formant frequency patterns that differentiate vowels and consonants are determined by the shape of the SVT. The formants of the vowel [i] necessitate a wide pharynx and constricted oral cavity, and those of the vowel [a] necessitate a constricted pharynx and wide oral cavity. However, the length of the SVT, which varies from person to person, results in different absolute values for the formant frequencies of the 'same' vowel or consonant produced by different individuals. Human listeners unconsciously take this into account by estimating the length of a speaker's SVT from the speech signal itself.

We, in effect, normalize formant frequencies, implicitly taking into account the anatomical constraints of speech production. The vowel [i], which cannot be produced by nonhuman SVTs that lack a 1:1 oral/pharynx proportion, is the optimal signal for this process. Because infants are able to normalize vowels between the ages of three and six months, this neural 'computation' seems to be innate. But the neural basis for normalization is another example of preadaptation; monkeys and many other animals (perhaps even alligators) seem to use formant frequencies to estimate the size of another animal. Preadaptation from this function to linguistic communication clearly occurred

and may have driven the restructuring of the human SVT to optimize the process.

The neural system of the human brain that regulates speech production is closely tied to the comprehension of language; its evolution again seems to involve preadaptation. The meaningful vocalizations of chimpanzees and monkeys often have different formant frequency patterns. However, the vocal signals of nonhuman primates are tied to specific emotions or affectual states, whereas humans produce arbitrary vocal signals that convey referential information. Humans are able to freely sequence the motor commands that generate vocal signals to form words. Although Broca's and Wernicke's areas of the cortex have been traditionally identified as the traditional language areas of the human brain, subcortical basal ganglia structures, which also sequence the motor commands necessary for walking and other aspects of motor control, are critical elements of a complex distributed neural system that confers cognitive flexibility, allowing humans to comprehend sentences with complex syntax and adapt thought processes to changing circumstances. Recent studies show that the basal ganglia perform similar sequencing operations, both in motor control and in cognitive and linguistic operations. Moreover, the neural structures that generate overt speech maintain words in short-term verbal working memory to comprehend the meaning of a sentence by the use of syntactic (arrangement of words), semantic (meaning), and pragmatic information. Because one function of the subcortical basal ganglia in humans, other mammals, and species ancestral to all mammals is to regulate adaptive motor control, preadaptation is evidently responsible for their role in the neural system that confers human linguistic ability.

References

Darwin, Charles. 1859. On the origin of species. London: Murray, Cambridge, MA: Harvard University Press, Facsimile edition, 1964.

Gibson, Katherine R., and Thomas Ingold. 1993. Tools and language in human evolution. Tools, language and cognition in human evolution. Cambridge: Cambridge University Press.

Hauser, Marc. D. 1996. The evolution of communication. Cambridge, MA: MIT Press.

Kimura, Doreen. 1993. Neuromotor mechanisms in human communication. Oxford, New York: Oxford University Press.

Lashley, Karl S. 1951. The problem of serial order in behavior. Cerebral mechanisms in behavior, ed. by L.A. Jefress, 112–46. New York: Wiley.

Lieberman, Philip. 1967. Intonation, perception and language. Cambridge, MA: MIT Press.

Lieberman, Philip. 1984. The biology and evolution of language. Cambridge, MA: Harvard University Press.

Lieberman, P. 2000. Human language and our reptilian brain: the subcortical bases of speech, syntax, and thought. Cambridge, MA: Harvard University Press.

Mayr, E. 1982. The growth of biological thought. Cambridge, MA: Harvard University Press.

Negus, Victor E. 1949. The comparative anatomy and physiology of the larynx. New York: Hafner.

PHILIP LIEBERMAN

See also **Anatomy of the Articulatory System; Evolution of Language 1: Overview; Evolution of Language 2: Cognitive Preadaptation; Evolution of Language 4: Social Preadaptations**

Evolution of Language 4: Social Preadaptations

'Selfish gene' Darwinism differs from earlier versions of evolutionary theory in its focus on one key question: Why cooperate? The faculty of speech is an aspect of human social competence. By inference, it evolved in the context of uniquely human strategies of social cooperation.

Noam Chomsky redirected linguists' attention away from social issues, defining language as an innate competence of individuals. To study the origins of language, however, we must strike out in new directions, reconnecting linguistics with Darwinism in its modern, socially aware form. Signal evolution theory states that where social conflict exists, low-cost cryptic communication gives way to costly and elaborate display aimed at overcoming the skeptical resistance of receivers. On this basis, human speech—which is low-cost, conventional communication—can only have emerged under very different social conditions, in which listeners could afford to vest trust in one another's signaling intentions.

Communication begins not when an individual makes a sign, but when someone else interprets that action as significant. Even a purely instrumental action, after all, may be socially understood as a signal. Where this has evolutionary significance, the behavior may then undergo modification in the service

of novel, socially conferred, signaling functions. If this is accepted, syntactical structure in the case of speech may have evolved by becoming progressively elicited and then consolidated by generations of comprehending listeners. First, conceptual complexity is 'read into' signaling by the attentive mind reader; subsequently, the signaler—given such encouragement—may succeed in externalizing aspects of that complexity explicitly in the signal itself.

In keeping with this scenario, one speculative hypothesis is that speech evolved thanks to novel levels of care, solicitude, and understanding shown by human caregivers toward their offspring. Chimpanzee infants may use a characteristic 'nursing poke' as a request for the mother's nipple. To reach the nipple, the infant at first simply pushes aside its mother's arm. As mother and infant cooperatively interact, the infant saves time and effort by relying on shorthand. The abbreviated version now bears some resemblance to a linguistic sign. A distance has appeared between the gesture's underlying meaning and its surface form—suggesting the rudiments of a 'secret code'.

Among apes, however, no such code survives into adulthood. To understand why ape codes fail to survive, we must examine the social context. Whereas the human infant may anticipate long-term kin-based solicitude, benefiting from social provisioning beyond infancy, a chimp youngster must fend for itself once it has been weaned. Deprived of the prospect of caring support, it abandons not only the now irrelevant nursing poke but also most other subtle indications of need. Given the competitive exigencies of impending adulthood, the best preparatory training for the ape youngster may in fact be to *avoid* excessive reliance on cooperative understanding from others. From this perspective, elaboration of symbolic potential as young apes mature appears constrained less by cognitive deficits than by a decisive *social* one—the obvious absence, in the wild, of any unconditionally supportive or caring audience. Why bother to elucidate one's aims or interests to others who may at best show indifference—or at worst exploit such intelligence for their own ends?

When delivered in sequence, ape vocal calls emphasize a single holistic meaning. If the calls vary, they may yield at most a blended compromise between meanings. By contrast, the playful bodily gestures of ape youngsters may be rich in cognitive information and combinatorial complexity. A gesture indicating 'Let's play!', for example, may systematically reverse the significance of subsequent 'chases' or 'bites', in ways reminiscent of a grammatical marker in language. Young primates frequently engage in make-believe play, whose inventiveness may almost suggest human verbal creativity. If we are seeking an evolu-

tionary precursor for human linguistic communication, the most convincing candidate is not the vocalizations of primates but their imaginative and often humorous social play.

Conversational speech including humor in the human case extends and develops the communicative potential of immature primate play. But if this is accepted, we must ask a new question. During the course of human evolution, how did the conditions for such creativity come to be extended from infancy into adulthood? The key restriction on animals' freedom to play is reproductive competition and conflict. In many species, the onset of sexual maturity brings with it the Darwinian imperative to engage in potentially lethal sexual competition. In the primate case, this impinges upon life concurrently with sexual maturity, setting up anxieties, conflicts, and resort to defensive alliances that effectively constitute adult sociality. As competitive stresses intensify, the tendency is for play fights to give way to real ones—whereupon the play stops.

Among humans, the transition to adulthood takes a different form. Youngsters go through an extended period of childhood, during which they can rely on social as opposed to 'fend-for-yourself' provisioning. At a certain point, young hunter–gatherer adolescents become coercively incorporated into ritual coalitions. Rites of initiation—central to intergenerational cultural transmission—may be viewed as spectacular 'pretend-play' performances, drawing on hallucinatory techniques such as trance, dance, rhythm, body-painting, and so forth. Whether or not genital mutilation is involved, the declared aim is to curb individualistic pursuit of sexual advantage. Bonds of coalitionary solidarity, typically modeled on sibling solidarity, are accorded primacy over sexual bonds.

The outcome of such solidarity is a domain of institutional reality—of objective facts whose existence depends, paradoxically, on subjective belief. Within human kinship systems, for example, a woman is 'our sister' (or a man 'our brother') because the collectivity asserts it to be so. When in America today, a piece of green paper counts as a dollar bill, that too is an institutional fact, as is the belief that 'bachelor' in English means 'unmarried man'. Children engaged in games of 'let's pretend' grasp the basic principle of institutional reality when they assert, 'this rag is mummy' or 'that stick is a horse'.

During a critical transition in human evolution, the institutional realities of cultural kinship acquired supremacy over purely biological reproduction. Collective regulation of reproductive relationships substantially reduced former levels of internal, sexual, and other conflicts. The effect was to open up a new social space in which the institutional facts of language could cumulatively evolve for the first time.

References

Dawkins, R. 1989. The selfish gene, 2nd edition. Oxford: Oxford University Press.

Deacon, T. 1997. The symbolic species: the co-evolution of language and the human brain. London: Penguin.

Dunbar, R., C. Knight, and C. Power (eds.) 1999. The evolution of culture. Edinburgh: Edinburgh University Press.

Hurford, J.R., M. Studdert-Kennedy, and C. Knight (eds.) 1998. Approaches to the evolution of language: social and cognitive bases. Cambridge: Cambridge University Press.

Knight, C. 1991. Blood relations. Menstruation and the origins of culture. New Haven and London: Yale University Press.

Knight, C., M. Studdert-Kennedy, and J.R. Hurford (eds.) 2000. The evolutionary emergence of language. Social function and the origins of linguistic form. Cambridge: Cambridge University Press.

Power, C., and L.C. Aiello. 1997. Female proto-symbolic strategies. Women in human evolution, ed. by L.D. Hager, 153–171. New York and London: Routledge.

Searle, J.R. 1996. The construction of social reality. London: Penguin.

Velichkovsky, B.M., and D.M. Rumbaugh (eds.) 1996. Communicating meaning. The evolution and development of language. Mahwah, NJ: Erlbaum.

Zahavi, A., and A. Zahavi. 1997. The handicap principle. a missing piece in Darwin's puzzle. New York and Oxford: Oxford University Press.

CHRIS KNIGHT

See also **Evolution of Language 1: Overview; Evolution of Language 2: Cognitive Preadaptations; Evolution of Language 3: Physical Preadaptations**

Ewe and Gbe languages

The Gbe languages are spoken in an area that begins in Lower Volta in Ghana and stretches through Togo and Benin to Western Nigeria and Lower Weme. These languages belong to the Kwa family, and have been treated variously as a dialect cluster and a language cluster. Earlier, Anglophones used to refer to the whole cluster as Ewe while Francophones preferred Aja, Adja, or Adja-Tado, which are names of major dialects/languages within the cluster. The term 'Gbe' was introduced by Hounkpati Capo to refer to the whole cluster because the majority of speakers designate their speech forms by adding *-gbe* to their ethnic names. Capo divides Gbe into five major dialect clusters comprising from West to East Ewe, Gen, Aja, Phla Phera, and Fon. According to Capo, 'there is mutual intelligibility between dialects that are contiguous, e.g. Ewe and Gen, Gen and Aja, Aja and Fon, etc.; but the degree of mutual intelligibility is related to geographical distance'.

Each of Capo's dialects constitutes a dialect cluster by itself, some of which are provided in Table 1.

TABLE 1 Gbe Dialect Cluster

Ewe		Gen	Aja	Phla-Phera	Fon
Adan	Kpele	Agoi	Dogbo	Alada	Agbome
Anfoe	Peki	Anexo	Hwe	Ayizo	Arohun
Anlo	Tonu	Gen	Sikpi	Kotafon	Kpase
Gbin	Ve	Gliji	Tado	Saxwe	Gun
Ho	Watsyi			Tofin	Maxi
Kpando	Notsie			Xwela	Weme

The Ewe dialects have been grouped into Coastal Ewe, which includes Anlo and Tonu, Central Ewe, which includes Ho and Kpedze, and Northern Ewe, which includes Anfoe and Kpando. The Central and Northern dialects are referred to as *Ewedomegbe*. While considerable variation exists within the Ewe cluster, the neighboring Gen cluster spoken in Togo and Benin is said to lack noticeable variations.

History

Oral traditions suggest that the Gbe people once lived in Ketu, a Yoruba town in present-day Republic of Benin. From there, a group moved south, and some of them settled at Tado near the Mono river. This settlement is associated with the Aja cluster. Some of those who moved south founded a settlement at Notsie (ŋɔ tsie), which is associated with Ewe. The Ewes in Ghana say their forefathers fled the tyranny of a chief called Agokoli and moved to settle in Ghana. A different group went from Ketu to settle in Adele in Togo for a while but later left to join the settlement in Notsie. Later, some of those who settled at Tado moved to found the Alada kingdom whose political nucleus was Agbome and Xɔgbonu. This settlement is associated with the Fon cluster. The Gen dialect is due to later settlement by the Fante-Aŋɛ from Elmina in Ghana who settled at Anexo, and the Ga also from Ghana who settled in the plains between Lake Gbaga and the Mono River.

Phonology

In all, 42 consonants and 16 vowels have been identified within the Gbe cluster. These are provided in Tables 2 and 3.

Not every consonant or vowel is used in every dialect. For example, Ewe is the only dialect cluster that has the voiceless and voiced bilabial fricatives, written [ƒ] and [ʋ], respectively. Some consonants also occur in complementary distribution. For instance, the nasal stops occur with nasalized vowels in Ewe while nonnasal variants occur with oral vowels, e.g. *nyɔ́* 'drip' (or *ỹɔ́* in some dialects). Also, the trill /r/ occurs only after dentals, alveolars, and palatals while the lateral occurs everywhere else, including syllable-initial position. With respect to vowels, Gɛn and Aja have merged [ɛ] and [e] into [e], while Fon and Phla Phera have retained the distinction. Within Ewe, the Coastal dialects have replaced [ɛ] with [ə].

Tone

Gbe is a tone language, i.e. intonational pitch is used to distinguish words. All the dialects use three tones namely, High, Mid, and Low. These three tones are subgrouped into High and Non-High. Consonants play an important role in the realization of these tones. Thus, a Non-High tone in a noun is realized as Mid if the noun root's consonant is a sonorant ([m, l,…]) or voiceless obstruent ([p, t, …]), as in *aŋē* 'plastic'. Where the consonant is a voiced obstruent ([b, d, g,…]), the tone is realized as Low. By contrast, a High tone in a noun is realized as High if the noun has a voiceless obstruent root consonant, and as Mid if it has a voiced obstruent. This Mid tone becomes low when the containing syllable occurs before a syllable with a high tone. In addition to these three tones that all the dialects possess, Anlo, a variant of Ewe, has an Extra High tone, and Adangbe has an Extra Low tone, which occurs on the last syllable in questions. This shows that while tone is distinctive and each lexical item possesses a basic tone, it is still highly variable in context. As such, it is not represented in the orthographic convention of the Gbe dialects. For a few words, the tones are written in Ewe to avoid confusion with other very similar words. These are the pronouns *è* 'you' vs. *e* 'he, she, it', *wò* 'you' vs. *wo* [plural marker], *nè* 'you' vs. *ne* 'if', and *lé* 'catch vs. *le* 'be located'. The combination of the three tones yields six contour tones namely, Mid–Low, High–Low, High–Mid, which are falling tones, and Low–Mid, Low–High, and Mid–High, which are falling tones.

Syllable Structure

The syllable in Gbe is predominantly open and, in the few cases where it is closed, the final consonant is always a nasal. Also, it only allows two syllable-initial consonant clusters, where the second one has to be a liquid ([l, r,…]) or an approximant ([y, w,…]), e.g. *klɔ́* 'wash', *tré* 'seal' *fiá* [fja] 'burn', and *bua* [bwa] 'pretend'.

Morphology

Nouns in Gbe can occur with the prefixes [a-], [o-], [e], [ə-], and ɛ-. While [a-] occurs in all the dialects and is obligatory for the words it occurs with, e.g. *agba* 'plate' (instead of just *gba*), the rest are optional dialectal variants (optionality indicated by parentheses). Thus, 'breast' is *(ɛ)no* 'breast' in Ho and *(ə)no* in Anlo. The Gbe dialects also have a productive reduplication process that derives different types of lexical items, including nominals, from a verbal base. Reduplication involves the copy of part of the verb root to the left. The part that is copied depends on the type of Gbe dialect. In Fon, only the initial consonant is copied, to which /i/ is added. Hence, *gba* 'build' becomes *gbigba* and *wlan* 'write' becomes *wiwlan.* In Ewe, the initial consonant together with the first vowel is copied; hence, *va* 'come' becomes *vava.* However, the second consonant in a syllable-initial cluster is not copied, e.g. *trɔ* 'turn' becomes *tɔtrɔ.*

Syntax

Gbe sentences have a basic Subject–Verb–Object word order. This order can be changed in Ewe through the focusing of an argument, e.g. *Amí kpɔ́ Kofi* 'Ami looked at Kofi' can become *Kofié Amí kpɔ́,* with *Kofi* being marked by the focus marker *-é* and moved to the sentence-initial position. The progressive and related constructions in most Gbe dialects use an auxiliary verb, e.g. *Amí le Kofi kpɔ́ḿ* 'Ami is looking at Kofi'. Note that the complement *Kofi* occurs in the preverbal position, while the progressive marker *-ḿ* is attached to the verb. In some variants, a simple high tone serves as the progressive marker. Some dialects like Aja-Dogbo and Kpesi do not allow the complement in the preverbal position.

Some verbs in Gbe are frequently mistaken for adjectives because they describe properties that are expressed by adjectives in other languages. However, they behave like other verbs in Ewe, e.g. *lolo* 'become big' takes the verbal habitual suffix *-na*. Gbe also has obligatory complement verbs (OCVs). These are verbs that cannot occur without complements. Some of the verbs are fully meaningful by themselves in the sense that they have translational equivalents in languages like English, but they must still occur with complements of generic meaning, e.g. *ɖu nú* 'lit. eat thing'. Some verbs with translational equivalents occur with complements that derive from the same root, e.g. *fi fi*

TABLE 2 Consonants in Gbe

	Bilabial	Labiodental	Lamino-dental	Laminopost-alveolar	Apical post-alveolar
Voiceless stops	p		t		
Voiced stops	b		d		ɖ
Nasal Stops	m				n
Voiceless affricates			tˢ		
Voiced affricates					
Voiceless Fricatives	ɸ	f		s	
Voiced Fricatives	β	v		z	
Tap					ɾ
Oral trills					r
Nasal trills					r̃
Oral laterals					l
Nasal laterals					l̃
Oral approx					
Nasal approx					

	Alveopalatal	Palatal	Velar	Uvular	Labio-palatal	Labio-velar
Voiceless stops			k			kp
Voiced stops			g			gb
Nasal		ɲ	ŋ			
Stops			ŋʷ			
Voiceless affricates		tʃ				
Voiced affricates		dʒ				
Voiceless Fricatives		ʃ		χ (χʷ)		
Voiced Fricatives		ʒ	ʁ (ʁʷ)			
Tap	l					
Oral trills						
Nasal trills						
Oral laterals						
Nasal laterals						
Oral approx		y	ɣ		ɥ	w
Nasal approx		ỹ			ɥ̃	w̃

TABLE 3 Vowels in Gbe

	Front	Central	Back
High	i, ĩ		u, ũ
Mid-high	e, ẽ		o, õ
Mid-low	ɛ, ɛ̃	ə, ə̃	ɔ, ɔ̃
Low		a, ã	

'lit. steal a steal'. A well-known type of OCV is the inherent complement verb (ICV), which is widely discussed in the literature on Kwa languages. These are verbs with indeterminate meanings that appear to be specified by their complements. The combination of an ICV and its complement usually yields concepts that are expressed by simple verbs in languages like English. An example is *fú*, which occurs with *du* 'race, course' to yield 'run', and *tsi* 'water' to yield 'swim'.

There are two types of double object verbs (DOV) in Gbe. There is the 'normal' type of DOV the only examples of which are *fíá* 'teach', and *ná* 'give', and there is the ICV type like *da* 'throw'. The complement position of the normal-type DOV is variable: either the Theme (the object transferred) or the Goal (the entity to which the transfer is effected) can occur immediately after the verb. Thus, both *fíá ha Kofi* and *fíá Kofi ha* mean 'teach Kofi a song'. The ICV-type DOV only allows the Theme to occur immediately after the verb; hence, *da kpé Kofi* 'throw a stone at Kofi'. There is a restriction on objects such that the Theme cannot be determined while the Goal is not. Thus, *Kofi ná ga lá ɖeví 'Kofi gave a child the money' is unacceptable.

Gbe also has a serial verb construction (SVC), which consists of two or more verb phrases in a clause without any syntactic marking of dependency. Additional properties of the clause include having one syntactic subject and one tense value. Also, no verb in the clause can be negated independently. An example of an SVC is *Kofi tsɔ ga lá fle agbalẽ* 'Kofi bought a book with the money.'

The Noun Phrase

The noun can occur in a clause without a determiner in which case it refers to a nondefinite entity. Thus, one can say *avu le afí má* 'lit. dog located there' to locate an animal belonging to the dog family. *Áɖé* 'a certain' is added when the speaker knows the entity. The definite article, which is *(l)á* in Ewe, is only used when the entity is known to the speaker and hearer. Gbe is also known for its logophoric pronoun. This is a pronoun used to refer to a person whose words, wishes, or thoughts are reported. The logophoric pronoun, which is *ye* in Ewe, can occur after every verb provided it occurs in a subordinate clause introduced by *bé* 'say, that', e.g. *Kofi yi be yeava kpɔ Ami* 'lit. Kofi went that he [Kofi] see Ami'.

Preposition and Postpositions

Gbe has prepositions that are derived from verbs and postpositions that are derived from nouns. The prepositions in Ewe constitute a closed class of about ten elements. They are distinguished from verbs by their inability to occur with the habitual suffix *–na*. For example, *tsó* is a verb in *Kofi tsóna le ga adé me* 'Kofi (usually) sets off at six o' clock', and a preposition in *Kofi mlɔna anyi tsó (NOT tsóna) ga ade me* 'Kofi usually sleeps from six o' clock'. Unlike Ewe, the other Gbe dialects do not have a habitual suffix and, therefore, it is difficult to determine which elements constitute a preposition without recourse to meaning. Postpositions are thus named because they occur after the noun and designate axial parts and regions of objects. Although they have mainly evolved from nouns, they are a different class of words. In Ewe, they are distinguished by their inability to take the possessive marker *ƒe*. Contrast *Kofi ƒé megbé* 'Kofi's back', where *megbé* is a nominal, with *Kofi megbé* 'behind Kofi' where it is a postposition. However, the postpositions in Gbe and other Kwa languages cannot be said to form a word class with prepositions, unlike e.g. in Japanese where the postpositions are true 'prepositions' that occur in postnominal position.

Ideophones

Finally, Gbe has ideophones — a set of words with interesting phonological and syntactic properties, e.g. *bɔhɔbɔhɔ* 'describing the heavy walk of a fat person', i.e. to dawdle. It also has utterance particles that signal the type of utterance or the attitude of the speaker, e.g. *sea* 'you hear?'

References

Aboh, Enoch. 2003. The morphology of complement-head syntax. New York: Oxford University Press.

Ameka, Felix. 1991. Ewe: Its grammatical constructions and illocutionary devices. Ph.D Dissertation. Canberra: Australian National University.

Ansre, Gilbert. 1966. The grammatical units of Ewe: a study of their structure, classes and Systems. Ph.D Dissertation. University of London.

Capo, Hounkpati B.C. 1984. Elements of Ewe-Gen-Aja-Fon dialectology. Peuples du Golfe du Benin (Aja-Ewé), ed. by François de Medeiros. Paris: Éditions Karthala.

Capo, Hounkpati B.C. 1991. A comparative phonology of Gbe. Publications African Languages and Linguistics, Vol. 14. Berlin and New York: Foris.

Clements, G.N. 1975. The logophoric pronoun in Ewe: its role is discourse. Journal of West African Languages 10. 141–77.

Duthie, Alan. 1996. Introducing Ewe linguistic patterns. Accra: Ghana Universities Press.

Duthie, Alan, and R.K. Vlaardingerbroek. 1981. Bibiography of Gbe (Ewe, Gen, Aja, Fon, Xwla, Gun, etc.). Basel: Basler Afrika Bibliographien.

Lefebvre, Claire, and Brousseau Anne-Marie. 2002. A grammar of Fongbe. Berlin and New York: Mouton de Gruyter.

Medeiros (de), François (ed.) 1984. Peuples du Golfe du Benin (Aja-Ewé). Paris: Éditions Karthala.

Pazzi, Roberto. 1980. L'homme Ewe, Aja, Gen, Fon et son univers. Lome: Nouveau Tirage.

Westermann, Diedrich. 1930. A study of the Ewe language. London, New York and Toronto: Oxford University Press.

JAMES ESSEGBEY

F

Fanakalo

Fanakalo is used in interethnic contacts in the eastern parts of South Africa. It is referred to variably as a pidgin with Zulu or Nguni vocabulary, Nguni being the Bantu subgroup to which both Zulu and Xhosa belong.

Although there is no definite consensus (there is a tendency to associate the language with the Indian emigration starting in the 1860s), it would appear that the history of Fanakalo starts with the contact between the British and speakers of Bantu languages in the Eastern Cape/Natal area from around 1840. As it proved to be a useful vehicle of interethnic communication, it soon received official encouragement—Fanakalo is rare among pidgins in that it has spread in part through formal education, as the South African Chamber of Commerce arranged courses in it in order to facilitate comprehension between miners of differing ethnic backgrounds. When European-run mining operations spread in the second half of the nineteenth century, so did Fanakalo. In the 1890s, the pidgin began to be used in Rhodesia and in the Copperbelt region of Zambia. It even ventured into the Katanga (Shaba) province of Belgian Congo, and its use has also been attested in Malawi and Namibia. In the beginning of the twentieth century, it was suggested that Fanakalo might even become *the* lingua franca of South Africa, and perhaps even of the entire southern parts of the continent.

The expansion was halted, however, and rapidly turned into a retreat. In Belgian Congo, an official decision was taken in 1918 to replace Fanakalo by Swahili as the working language, and in Zambia, the somewhat streamlined variety of Bemba known as Town Bemba gradually ousted Fanakalo. Problems gathered even on its own home turf, as Fanakalo became increasingly perceived as a means of oppressing the Black proletariat, and barring its members from learning English.

Fanakalo continues to be used in South Africa, despite explicit efforts on behalf of authorities to stamp it out. In 1995, for instance, an official reported that English ought to be used in the mines instead of Fanakalo, pretending that the latter was not an 'adequate medium of communication to ensure underground safety'. Although still known by hundreds of thousands of Africans, Fanakalo has lost much of its raison d'être, and it is most likely doomed unless the use of English in South Africa ceases to increase.

More than two thirds of the lexicon of Fanakalo is Bantu in origin—mostly Zulu, but influences from Xhosa have also been pointed out. Slightly less than a fourth of the vocabulary is derived from English, while Afrikaans has contributed the remainder. In most basic vocabulary, not unexpectedly, Zulu is dominant. On the 200-item 'Swadesh list' of basic vocabulary, only eight words are not of Zulu origin. There are some doublets in the lexicon, as in the numeral system, where English and Zulu items are in free variation.

In the segmental phonology, the most obvious departure from local Bantu languages is the substitution of plosives for clicks — at least among native speakers of nonclick languages. The implosives of Zulu have become explosive in the pidgin, but two lateral fricatives are nevertheless retained. The European component of the lexicon, on the other hand, has been largely adapted to the requirements of Bantu syllable structure. Thus, Afrikaans *sterk* 'strong' surfaces as *stelek* in the pidgin.

While Zulu is a tone language, Fanakalo does not use pitch differences to distinguish words. Also, vowel length and consonant aspiration are irrelevant. Intonational stress is on the penultimate syllable, as in Zulu. The vowel system includes five phonemes, as in the local Bantu languages.

Fanakalo is a strict S(ubject)–V(erb)–O(bject) language, as are most of those on which it is based. It is noteworthy that adjectives and numerals precede the modified nouns, as in English and Afrikaans, but contrary to Zulu and its relatives.

Although most meanings are expressed by independent words, Fanakalo does have a few bound morphemes, all of which are inherited from Zulu: past, passive, and causative are all marked by suffixes (*-ile*, *-wa*, and *-isa*, respectively).

More in line with what one would expect from a pidgin, the intricate Bantu agreement system has left few traces in Fanakalo. However, different observers disagree on the fate of the Bantu noun class system. It would seem unjustified to describe Fanakalo as a class language, although the plural prefix *ma-* has variants that distinguish whether the noun refers to animate or inanimate entities. However, Fanakalo only marginally preserves other features of the original noun class distinctions.

Different from Zulu, the future morpheme *zo* is unbound (i.e. not an affix), as is the determiner *lo*, which is perceived as shibboleth of the language to the extent that one of its aliases is *isilolo* — literally 'the lo-lo language'. The sentence negator is a free morpheme, and the bound pronominal markers of Bantu languages correspond to free morphemes in Fanakalo. These pronouns demonstrate no case distinctions, and many varieties simplify even further through the use of 2sg and 3sg plus *zonke* 'all' for the corresponding plural forms.

Clauses generally do not use a copula ('to be'), although *kona* 'to have' has a secondary use as existential copula.

Both conjunctions and subjunctions are attested, but subordination is normally replaced by coordinating strategies, often simple juxtaposition of sentences.

Several prepositions are also found, although *lapa* is semigeneralized in that it covers a wider range of uses than most prepositions in the input languages. Also common is *ga*, which — placed between the possessum and the possessor — marks genitival relationships.

Few details are known on the variation of Fanakalo. Lexical differences have been observed, however, between the different countries where pidgin is used. Not surprisingly, the Zimbabwean variety is reported to display influences from Shona, the major language of the country, in addition to having a larger English and a smaller Afrikaans component.

References

Adendorff, Ralph. 1993. Ethnographic evidence of the social meaning of Fanakalo in South Africa. Journal of Pidgin and Creole Languages 8(1). 1–27.

Aubry, Caroline. 1997. Kitchen kaffir, basic Bantu, isilololo, fanagalo... étude ethnolinguistique d'un pidgin sud-africain aux multiples visages. M.A. Thesis. Département d'anthropologie, Université de Montréal.

Cole, D.T. 1953. Fanagalo and the Bantu languages in South Africa. African Studies 12(1). 1–9.

Ferraz, Luís Ivens. 1980. Notes on a pidgin dialect. African Studies 39(2). 209–20.

Heine, Bernd. 1973. Pidgin-Sprachen im Bantu-Bereich. Berlin: Dietrich Reimer.

MIKAEL PARKVALL

See also **Pidgins and Creoles; South Africa; Zulu and Southern Bantu Languages**

Farsi

Farsi is one of the three major variants of Persian, a member of the Iranian language family. The other two variants are Dari, spoken in Afghanistan, and Tajiki, spoken in Tajikistan. The Iranian languages belong to the Indo-Iranian branch of the Indo-European language family. Farsi is the official language in Iran, and is the native tongue of about half the population of this country. Other Iranian languages spoken in Iran include Gilaki, Kurdi, Lori, and Baluchi.

The evolution of Persian as the culturally dominant language of the Near East began with the political domination of the dynasties that originated in the province Pars. The name of the language, Parsi, is associated with this ancient province. After the Arab invasion in the sixth century, Parsi was utilized as *Farsi* in official documents written in Arabic due to the lack of /p/ in this language. *Farsi* is now the term used for the official language, while *Parsi* is used mostly in elevated literary writings.

Old Persian

The oldest version of Farsi is known as Old Persian, the official language of the Achæmenian Dynasty who had originated in the province Pars. Evidence of this ancient language is found in Achæmenian inscriptions, tablets, and coins. There is no document available of this language following the fall of the Achæmenian Dynasty in the third century BC.

Old Persian was structurally similar to Latin and Ancient Greek. It was a Subject–Object–Verb language with a fairly free word order in the preverbal position. The case system was rich, consisting of seven cases realized as suffixes. Here is a sample:

Nominative:	martiya 'man'
Accusative:	martiyam
Vocative:	martiyaa
Instrumental:	martiyaa
Locative:	martiyai
Genitive:	martiyahyaa
Ablative:	martiyaa

The language was basically synthetic, and the number of functional words was very limited.

Nouns and adjectives were conjugated for masculine, feminine, and neuter, and the number system consisted of singular, dual, and plural. The copula was often lacking, as illustrated by the following example:

> *Baq vazrek Ahuramazdaa*
> God great Ahuramazda
> 'The great God (is) Ahuramazda.'

Middle Persian

The oldest document available in Middle Persian, the descendent of Old Persian, belongs to the first century. That is, there is a gap of four centuries between the last documents found in Old Persian (third century BC) and the first one discovered in Middle Persian.

Middle Persian consisted of several dialects and variants. One of these variants, known as *Parti*, *Pahlavani*, or *Pahlavi*, was the language of the Ashkanian Dynasty who governed in the northeastern part of Persia from the third century BC through the third century AD. Documents related to this variant are limited to a few short inscriptions and coins attributed to this dynasty, and a few inscriptions related to the era following their fall. Another variant of Middle Persian, called *Parsik* in some documents and *Pahlavi* in some others, was the official language of the Sassanian Dynasty, a family who originated in the province Pars, and was the last dynasty before Arab invasion in the sixth century. Most scholars have this variant in mind when using the term 'Middle Persian'. The last documents detected in this language belong to the tenth century, four centuries after the fall of the Sassanian Dynasty, at a time when it had already ceased to exist as a language in use.

Although the gap between the last document found in Old Persian and the first one discovered in Middle Persian does not exceed four centuries, the morphological structure of the latter had radically changed compared to the former. The final syllables in nouns and adjectives were lost. Consequently, the rich case system had disappeared. The number system was reduced to singular and plural, and there was no sign of the three-way gender system. The plural suffix *ân* was dominantly used, and the suffix *ihâ*, the ancestor of the generic plural suffix *hâ* in Modern Farsi, was restricted to religious books such as *Bondahishn* and *Minu-ye Kherad*. Unlike Old Persian, there were many functional words present in this language. In spite of the loss of the case system, the word order was still fairly free in the preverbal position. Also, the copula continued to be absent in most cases, as in *to ki* (*hi*) (you who (are)) 'who are you'.

Modern Persian

The oldest documents written in Modern Persian can be traced to the early tenth century. Persian soon became the language of world class poetry, and has undergone some minor morphological, phonological, and syntactic changes in the last 11 centuries. Of the three variants of this language, Tajiki has preserved the syntactic properties of Middle Persian more than the other two, while spoken Farsi reveals the greatest changes with respect to word order and other syntactic properties.

Properties of Farsi

The phonological system of Farsi consisted of eight vowels in its early stages, six of them short and two long. The long vowels did not survive, and the disappearance of the feature [long] in the phonological system created a number of homophones in this language (e.g. *sheer* and *shir*, both pronounced as *shir* now, meaning 'milk' and 'lion'). All remaining vowels are tense. There are 21 consonants and one semivowel /y/ in this system. Unlike Middle Persian, Farsi does not allow initial consonant cluster, although final consonant clusters are very common in this language.

The lexicon of Farsi has been highly influenced by borrowings from Arabic. It has been estimated that 65% of the vocabulary used in modern novels and short stories is of Arabic origin. This number increases when taking into account the conservative and formal writings. The influence of modern European languages such as French, and most recently English, is also evident in the lexicon of this language.

The morphological system consists of verbal prefixes and suffixes, the generic plural suffix *hâ*, the plural suffix *ân* (only for animate nouns), and derivational affixes utilized for nouns. Verbs are inflected for person and number. There is no gender distinction, not even in the pronominal system.

One of the interesting characteristics of Farsi is its verbal constructions. Middle Persian did not have many complex verbs, and the only light verb used in verbal constructions was *kardan* 'to do'. Since the thirteenth century, complex verbs have gradually replaced simple verbs. As a result, the number of simple verbs does not exceed 115 in contemporary Farsi; many of them are not even used in the everyday application of the language. The tendency of forming complex verbs has resulted in the existence of two sets of verbs, simple and complex, for a number of verbal concepts. In many cases, the application of the simple verb is restricted to the written and elevated language (e.g. *geristan*, *gerye kardan* 'to cry'). The productivity of complex verbs is evident by the fact that its nonverbal element is not restricted to native Persian elements, nor to Arabic infinitives that entered the language centuries ago, but it also includes recent borrowings from European languages (e.g. *telefon kardan* 'to call', *tâyp kardan* 'to type'). The verbal element of complex verbs is not restricted to the verb *kardan* 'to do' as it was in Middle Persian, but it ranges over a number of simple verbs that are used as 'light verbs' in those constructions (e.g. *zadan* 'to hit', *bordan* 'to take', *dâshtan* 'to have', etc.). Another characteristic of the complex verb in Farsi is that its nonverbal element ranges over a number of categories such as nouns (*shekast dâdan* (defeat giving) 'to defeat'), adjectives (*bidâr shodan* (awake becoming) 'to wake up'), adverbs (*bâlâ bordan* (up taking) 'to promote'), and even prepositional phrases (*az yâd bordan* (of memory taking) 'to forget'). These complex verbs receive either idiomatic reading (e.g. *dast andâkhtan* (hand throwing) 'to mock') or compositional interpretation (e.g. *fekr kardan* (thought doing) 'to think').

Similar to Middle Persian, Farsi lacks a case system. The only element arguably associated with case is *râ*, a particle that marks specific direct objects for accusative case, and has evolved from Old Persian *râdi* with the meaning 'for the sake of'. The Middle Persian reflex of *râdi* is *rây*, which serves several functions such as illustrating purpose or reference, and marking the indirect object. In late Middle Persian texts, *rây* starts taking on a new function by marking the direct object as in *shaw sheerân rây be-ozan* 'Go, slay the lions'. However, even in early stages of Modern Persian the usage of this particle following specific direct objects is not widespread as in *khosro Tus be d-u dâd* 'The King gave Tus to him.' Marking specific objects for accusative case has become the dominant function of *râ* only after the fifteenth century. Lack of *râ* still persists in some unofficial dialects, such as in Jahromi, one of the Southern dialects of Persian, as in *dâr-am ketâb-et mi-khân-am* 'I am reading your book'. It has been argued that some of the Middle Persian texts were rewritten in the sixteenth and seventeenth centuries, and thus the existence of *rây* as a direct object marker in those texts might be the influence of Modern Persian, the native language of the authors. In more recent literature, the primary function of *râ* has been suggested to be marking specificity rather than accusative case. The same idea has been proposed for similar cases in regional languages, for example, *ko* in Urdu and Hindi and possibly *-I* in Turkish.

One of the major syntactic differences between Middle Persian and Farsi is the lack of ergative constructions in the latter. In Middle Persian, the transitive verb in the past tense agrees with the object, as in *man to did hi* (I you saw-2sg) 'I have seen you.' Although the verb agrees with the subject in Farsi, agreement between the nonanimate subject and the verb is optional, as in *barg-hâ sabz shod* (*and*) 'the leaves became-sg (pl) green'. However, lack of agreement is possible only in an unmarked word order in these cases. Otherwise, agreement becomes obligatory, as in *sabz shod-and / *shod barg-hâ.* Another difference between Middle Persian and Farsi is that the occurrence of the copula has become obligatory in the latter: compare *to ki* (*hi*) (who you are) in Middle Persian with *to ki *(hasti)* in Farsi.

Ezafe construction is one of the specific syntactic properties of Farsi. It is a noun phrase consisting of the head (an element with the feature [+N] such as Noun or Adjective), its modifier(s), an optional possesser, and the Ezafe particle *e* that is structurally utilized as a link between the head and its modifier, (e.g. *dokhtar-e bâhush-e dust-e mâ* 'the smart daughter of our friend'.

The basic word order is S(ubject)–O(bject)–I(ndirect) O(bject)–Verb if the direct object is followed by *râ* and receives a specific reading. The nonspecific object is semantically closer to the verb, and is syntactically adjacent to it. Thus, the basic word order is S–IO–O–V when the object is nonspecific. Farsi is a Null Subject language, and thus lacks dummy subjects comparable to *there* and *it* in English. This language has maintained the fairly free word order observed in Old and Middle Persian, and therefore is classified as a *scrambling* language. Recent studies on scrambling languages indicate that the word order in these languages is not 'free', and that the rearrangements of phrasal categories are triggered by discourse functional features such as Topic and Focus or grammatical features such as the one representing the Extended Projection Principle (EPP). The

word-order variations in these languages are subject to the general principles of Universal Grammar, and may have an effect on the semantic interpretation of the linguistic expression. Thus, the term *free word order* does not reflect the syntactic facts in Farsi and similar languages. Finally, there is no structural wh-movement in this language, although wh-phrases are subject to scrambling similar to other phrasal categories.

References

Ghomeshi, Jila. 1997. Non-projecting nouns and the Ezafe construction in Persian. Natural Language and Linguistic Theory 15. 729–88.

Karimi, Simin. 1990. Obliqueness, specificity, and discourse functions. Linguistic Analysis 20. 139–91.
———. 1997. Persian complex verbs: idiomatic or compositional. Lexicology 3(2). 273–318.
———. 2003. On object positions, specificity, and scrambling, ed. by S. Karimi. Cambridge: Blackwell Publishers.
Khanlari, Parviz. 1976. Dastur-e Zab,n-e Farsi (The grammar of Persian language). Tehran: Bonyad-e Farhang.
Lazard, Gilbert. 1992. A grammar of contemporary Persian. California: Mazda Publishers.
Windfuhr, Gernot. 1979. Persian grammar: history and state of its study. The Hague: Mouton.

SIMIN KARIMI

See also **Indo-European 3: Indo-Iranian Languages**

Feature Theory

Classical Phonetics felt the need for a classification system for speech segments to give some indication of how sounds are made and at the same time to identify relationships between them. It used the place–manner system for consonants and the high–low/front–back system for vowels. The main purpose here was clearly to enable the phonetician to specify how particular sounds were made with respect to their articulation. There was, however, an important spin-off: it became possible to use the features or parameters of the classification system to label whole sets of sounds or articulations. Thus, we might refer to 'the set of all plosives' (seven in English: *p, t, k, b, d, g*, and the *glottal stop*), 'the set of all voiced plosives' (three in English: *b, d, g*), or 'the set of all voiced alveolar plosives' (one only in English: *d*), and so on, cutting horizontally and vertically around the consonant matrix. Similarly, for vowels, 'the set of all front vowels' (English: *i, e, æ*) 'the set of all rounded vowels' (English: *u, o*), and so on.

For the most part, the categories or features used by Classical Phonetics reflected what phoneticians regarded as the salient parameters of articulation: place features such as bilabial, dental, alveolar, palatal, velar, pharyngeal, laryngeal, etc., and manner features such as plosive, fricative, continuant, etc. This is not an exhaustive list, but the idea was to characterize for each consonantal sound *where* in the vocal tract the sound was made and *how* it was made. In the case of vowels, the relative position of the relevant part of the tongue on the high–low and front–back axes of a two-dimensional coordinate system within the oral cavity would be sufficient, using front–back features such as front, central, and back, and high–low features such as high, mid, and low.

It was sometimes necessary to add to these basic classificatory features to reflect details of the articulation, often depending on how finely the phonetician needed to specify articulations. In the case of vowels, some additional features were necessary to expand on the two-dimensional coordinate systems: vowels could in addition be spread or round, nasal or nasalized, and long or short; and consonants could be either tense or lax. For example, in English the high front vowel [i] is said to be long, whereas the low front vowel [æ] is said to be short; [t] and [s] are tense sounds, but [b] and [v] are not tense.

As a spin-off of being able to label sets of sounds or articulations in this way, it became possible to describe the *behavior* of various sets—that is, the behavior common to members of now identifiable sets could be characterized. Thus, for example, it was possible to say that the set of voiced plosives devoice in word-final position or that all vowels lengthen before voiced plosives in the same syllable, and so on. Hence, rules no longer had to be about the contextual behavior of individual sounds, but in terms of how sets or classes of sounds behave. Phoneticians now had the ability to capture and express generalization—an important theoretical principle in linguistics: generalizations *must* be expressed whenever possible. Capturing generalizations is the basis of predictive modeling because it means that we can ask the question: What about any other segment that has these same features?—even though we may

not yet have analyzed it. Generalization leads to predictions and hypotheses, and these form a strong basis for scientific advance.

For all its potential as the basis of a truly scientific study of speech, the feature classification system afforded by Classical Phonetics, however, rested in fact on very weak theory. For example, a given sound might have its relevant features identified, but phoneticians did not yet have a way of saying why there were gaps in how features could be combined, nor was it possible to predict why certain sounds could not in fact ever occur. Why can we not have a bilabial plosive that is also a fricative? Why is it not possible to have a voiced glottal fricative? We can easily supply answers to these questions—but it is hard to construct a theory around this system that predicts all the constraints and at the same time all the possibilities in an *explanatory* way—which is ultimately what we are after. We want to know *why* things are the way they are, as well as simply being able to characterize them *as* they are.

It was not until Transformational Generative Grammar came along, though, that these generalizations became formalized in recent phonological theory. Morris Halle's *Sound pattern of Russian* (1959) was really the first influential textbook in modern phonological theory (just two years after Noam Chomsky's *Syntactic structures* (1957), the first influential textbook in modern syntactic theory). The Generative Phonologists adopted the theory of distinctive features from the earlier Prague School of Linguistics (Nikolai Trubetskoy 1958 [1939])—a much more formal representational system than that of the classical phoneticians.

Distinctive Feature Theory in Phonology

The use of distinctive features in phonology enables us to capture *natural classes*, and, by extension, to generalize regularly occurring phenomena and to formulate predictions about the behavior of class members. If we wanted to hypothesize about human processing of phonology, we would use this idea to suggest that human beings process the patterns of phonology as part of speech planning in terms of these classes rather than in terms of individual segments. The regularity of patterning in phonology is part of the evidence for this claim—but the claim is more solid when based on the evidence that when the users of a language make up new words they do so by producing utterances that obey the rules of the natural classes their sounds fall into.

There have been various sets of distinctive features proposed as the parameters of segment description and classification. The original set appeared in Jakobson et al. (1952), and consisted of around 14 features; Jakobson and Halle (1956) had 12. Chomsky and

Halle (1968) had around 45 features, explaining that they found the original set of 14 somewhat inappropriate for characterizing some subtleties in phonology.

Most modern phonologists argue for a binary system of indexing features: a segment either possesses or does not possess any one particular feature. The point here is that whereas when describing the physical world of phonetics it may be useful to have a system capable of capturing multivalued features (an *n*-ary system), in the cognitive or perceptual world of phonology the binary system is preferable. For example, although we might find that there is somewhat more coarticulation between [−nasal] vowels and surrounding [+nasal] consonants (in words like *man* or *moon*) in some US accents of English than in most British accents of English, phonologically this observation is not relevant. There are no nasal vowels in English, although we may want to note that in some accents they are nasalized in some environments, the degree of nasalization being of no consequence.

Clearly, with a binary system of indexing, the maximum number of features needed to uniquely classify the sounds of a language like English (with around 45 phonemes) would be six, giving us 2^6 or 64 segments. More would be needed to uniquely classify the sounds of all the languages of the world or indeed all possible human languages. Larger sets of features were chosen because it was felt appropriate to sacrifice mathematical simplicity or elegance in favor of a feature-labeling system that appeared to relate these phonological features with the phonetic set of Classical Phonetics. Thus, the *meaning* of the features became more transparent.

These ideas are embodied in three principles surrounding the distinctive feature set. They should be able to

(1) characterize all *contrasting segments* in human languages;
(2) capture *natural classes* in a clear fashion; and
(3) be transparent with regard to *phonetic correlates*.

A claim inherent in the *first principle* is that the feature set might somehow embody the universal humanness of features rather than any language-*specific* nature. It is predicted that if this set is correctly specified, no other features will be needed even for future languages, as long as human beings do not change how they make and handle language—that is, as long as human beings remain human.

The *second principle* refers to not just classes, but *natural* classes. The idea here is that the classes themselves reveal something of what is natural in human language behavior, once again referring to the fact that phonological processing is a human activity, and will therefore contain elements that are truly universal.

The *third principle* enables us to establish phonetic similarity—that is, to group sounds that are phonetically similar by feature. In the end, there is a very good reason for doing this: it becomes possible to explain some phonological processes in terms of the behavior of their phonetic correlates.

The distinctive feature set most usually found these days is approximately that of Halle and Clements (1983), which is based on the Chomsky and Halle (1968) set. A detailed description can be found in Gussenhoven and Jacobs (1998). Chomsky and Halle have a lengthy description of their own set.

Feature Sets

The 12 features found in Jakobson and Halle (1956) can be related to what it is they characterize articulatorily and acoustically. These are abstract, phonological features; hence, their relationship with phonetic characterizations is one of relative correlation, rather than absolute definition. The physical correlates are only approximate—there are many variants on these simple descriptions both between languages and dialects, and between different speakers (see Table 1).

The Chomsky and Halle feature set is more comprehensive. There are some 27 basic articulatory features, although each has particular acoustic correlates. One point about these features is that they can be used at an abstract phonological or perceptual level, in which case they take on binary values, or they can be used in a fairly limited way at a physical phonetic level, in which case they can be multivalued. It is important to realize, though, that the correlation is not necessarily linear or one-to-one (see Table 2).

Example Use of Distinctive Feature Theory: Redundancy

Redundancy is an important aspect of phonology that is captured by the use of distinctive features. Consider, for example, the fact that all segments in English that are [+nasal] are also [+voice]. We could say that to specify [+voice] for segments like [m] and [n] is to fail to capture this redundancy. The main distinctive feature here is the nasality—the voicing is secondary and entirely predictable: all nasal consonants are voiced—remember we are discussing abstract phonology, not phonetics.

One of the principles involved in Distinctive Feature Theory is to set up a system to capture all the segmental contrasts in the world's languages. This can be done, and we can also show where there is no contrast: there is no contrast, nor possibility of contrast, where there is redundancy. If nasals are always voiced, then there cannot be a contrast involving voiceless nasals. Two things follow from this in the way we use features in the theory:

- we need only indicate those feature markings that contribute to the contrasts in a particular language;
- we can capture the redundancies in a separate table of metarules—rules outside the Distinctive Feature specification.

TABLE 1

Feature	Opposing Feature	Articulatory Correlation	Acoustic Correlation
Vocalic	Nonvocalic	Vocal cord vibration, relatively unobstructed vocal tract	Periodic vocal cord excitation and clear formant structure
Consonantal	Non-consonantal	Partial or complete vocal tract constriction	Overall energy relatively low
Compact	Diffuse	Front 'resonance chamber' dominates	Energy focused toward the center of the spectrum
Tense	Lax	Vocal organs relatively tense or the entire tract voluntarily greatly distorted	High energy, spread throughout the spectrum
Voiced	Voiceless	Vocal cord vibration present	Periodic (laryngeal) excitation
Nasal	Oral	Nasal cavity brought into play—the velum is lowered	Additional nasal formant(s) present
Discontinuous	Continuous	Vocal tract rapidly closing and opening	Interruption of the acoustic 'flow'
Strident	Mellow	Turbulence created at the place of articulation	Temporally unstructured noise at a relatively high frequency
Checked	Unchecked	Glottalized	Abrupt energy onset/offset
Grave	Acute	Marginal within the vocal tract (not central)	Energy focused on the lower part of the spectrum
Flat	Plain	Constricted aperture (example is lip-rounding)	Upper frequencies attenuated
Sharp	Plain	Constriction of the upper oral cavity, but relative widening or the lower cavity or pharynx	Lower frequencies attenuated

This table is based on Appendix 2 of Clark and Yallop (1990)

TABLE 2

Feature	Articulatory Correlation
Major class features	
1. Sonorant	Vocal cord vibration possible and usually present
2. Vocalic (or syllabic)	Vocal cord vibration possible, but vocal tract constriction restricted to vowel positions
3. Consonantal	Significant constriction present in the vocal tract Cavity features
4. Coronal	Blade of the tongue raised
5. Anterior	Constriction in front of the palatoalveolar place
6. High	Raised tongue body
7. Low	Lowered tongue body
8. Back	Retracted tongue body
9. Rounded	Labial rounding
10. Distributed	Extended place of articulation
11. Covered	Narrow and tense pharynx or raised larynx
12. Glottal constriction	Constriction produced by the vocal cords
13. Nasal	Velar port open
14. Lateral	Tongue sides lowered
Manner of articulation features	
15. Continuant ([−continuant]=stop)	Relatively unimpeded airflow
16. Instantaneous release	Sudden release (as in plosives)
17. Velar suction	Velar closure (used in clicks, for example)
18. Implosive	Glottal closure produces inward airflow
19. Velar pressure	(Unclear)
20. Ejective	Glottal closure and increased pressure
21. Tense ([−tense]=lax)	Musculature is highly contracted
Source features	
22. Raised subglottal pressure	Increased muscular contraction to raise subglottal pressure
23. Voiced ([−voiced]=voiceless)	The vocal cords are vibrating
24. Strident	Airflow turbulence produced at point of articulation
Prosodic features	
25. Stressed	(Unclear)
26. Pitch (relative, scalar)	(Unclear)
27. Length (relative, scalar)	(Unclear)

This table is based on Appendix 2 of Clark and Yallop (1990)

Omitting feature markings where there is redundancy means literally leaving the redundant cells blank in the distinctive feature matrix formed when segments have their feature specification characterized. The fact of the redundancy is captured by separate rules that take the general form:

if X then Y
or, in our specific example,
if [+nasal] then [+voice].

But why would we want to capture this redundancy, except to show that it is a regularity in the way segmental features pattern? We want to do this because speakers of a language *know* about the redundancy. Let us look at an example: there are three nasals in English: the nasal alveolar stop [n], the bilabial nasal stop [m], and the velar nasal [ŋ]. If we ask an English speaker to *invent* a new nasal—say, a palatal nasal like the one found in the French word *angeau*—they will also automatically make it [+voice]. It is as though they know that nasals must be voiced, which is another way of stating the above rule.

A very early original important text on this point was Richard Stanley's *Redundancy rules in phonology*, published in the journal *Language* in 1967 (Vol. 43). These redundancy rules were called *segment structure rules* to contrast them with another type: *sequence structure rules*. The latter capture a speaker's knowledge of redundancy in the specification of segments themselves in patterned sequences. Thus, if we have the sequence at the beginning of a syllable in English: CCCV... (with C=consonant, V=vowel), then the first C must be [s]—a completely redundant situation, since all we need to know is that there is a consonant there, followed by two others. In fact, there are heavy constraints also, of course, on the remaining two consonants: the second one must be a plosive and the third must be a liquid ([r, l]) of some sort or a semivowel ([y, w]). Some phonologists have pointed out that the onset of a syllable strings consonants together with increasing sonority until you get to the vowel nucleus (the supreme sonorous segment), followed by a coda of consonants of decreasing sonority—although there are exceptions to this principle.

Stanley's work, which like much of the earlier work in modern phonology, adopted an essentially *linear*

approach to analysis. Segments were characterized in terms of how they behaved with respect to their neighboring segments in the surface string, which represented the underlying phonological makeup of utterances. These days it is better to reconstruct such analysis in a *nonlinear* or hierarchical way, since this is much more revealing of the underlying constraints on the sequencing of segments on the surface. Nevertheless, whatever the approach the focus is on structure in phonology and the way such structure is particularly well expressed using sets of constraints on what can and cannot happen in a phonological system.

Explanation is the ultimate goal, and for this to be fully effective it must be sourced from outside the domain of linguistics itself. Thus, for example, we could use Distinctive Feature Theory to express an observed regularity that in many languages utterance final obstruents that are [+voice] often become [−voice], and this regularity is *explained* by reference to the failure of vocal cord vibration as the transglottal airflow decreases with falling subglottal air pressure—that is, the vocal cord vibration tends to stop as we run out of breath toward the end of an utterance. The explanation for the phonological regularity captured by regularities expressed through Distinctive Feature Theory lies *outside* phonology—in the domain of aerodynamics.

Distinctive Feature Theory makes a distinction then between the use of features for characterizing the contrastive properties of phonological segments, and using them to indicate redundancy. An *incomplete* distinctive feature matrix uses blanks to indicate redundancy (and lets you know where cells are the subject of redundancy rules), whereas a *fully specified* distinctive feature matrix has all cells filled with either a + or a − to indicate exhaustively just which features are present or absent for this particular segment.

Distinctive Feature Theory has been a significant step forward in classification from the rather crude phonetically based ideas of Classical Phonetics. Remembering, however, that it is essentially a concept in abstract phonology (rather than phonetics), its principal importance lies in how it lends itself to capturing the generality of phonological processes and the principles underlying the structure of phonological segments.

References

Chomsky, Noam. 1957. Syntactic structures. The Hague: Mouton.

Chomsky, Noam, and Morris Halle. 1968. The sound pattern of English. Cambridge, MA: MIT Press.

Clark, John, and Colin Yallop. 1990. An introduction to phonetics and phonology. Oxford: Blackwell.

Gussenhoven, Carlos, and Haike Jacobs. 1998. Understanding phonology. London: Arnold.

Halle, Morris. 1959. The sound pattern of Russian. The Hague: Mouton.

Halle, Morris, and George N. Clements. 1983. Problem book in phonology. Cambridge, MA: MIT Press.

Jakobson, Roman, and Morris Halle. 1956. Fundamentals of language. The Hague: Mouton.

Jakobson, Roman, Gunnar Fant, and Morris Halle. 1952. Preliminaries to speech analysis: the distinctive features and their correlates. Cambridge, MA: MIT Press.

Stanley, Richard. 1967. Redundancy rules in phonology. Language 43.

Trubetskoy, Nikolai. S. 1958. Grundzüge der Phonologie. Göttingen: Vandenhoeck and Ruprecht (reprinted from Travaux du Cercle Linguistique de Prague, 1939.

MARK TATHAM

See also **Halle, Morris; phonetics; phonology**

Ferguson, Charles Albert

Charles Albert Ferguson is best known for his 1959 article 'Diglossia'. In it he undertook to describe a particular 'language situation' in which two closely related but substantially different varieties of a language (one designated 'High', the other 'Low') exist side by side in a single speech community in a stable and functionally complementary relationship. This article has spawned approximately 3,000 studies of the topic, with no apparent end in sight. Some view its publication as the beginning of the field of sociolinguistics.

It has been reprinted in various collections and translated into a variety of languages.

Ferguson's interests were exceptionally broad, both geographically and topically. He spent a good deal of his career focused on Arabic and with languages of South Asia, but also wrote on speech communities and topics as varied as the phonology of Philadelphia, literacy in Diyari (an Australian language), loss of agreement patterns in Swedish, and the Ethiopian educational system. His interests bridged the 'applied'

and 'theoretical' divide. During much of the first half of his career, he was professionally occupied with solving practical language problems, such as developing language training programs or assisting developing nations in matters of language planning and policy. He spent the last half of his career at Stanford University, where he was able to pursue more theoretical research (e.g. child phonology and language universals), but he remained actively involved in applied linguistics.

Over the course of his career, Ferguson pushed for a far more comprehensive theory of language than was typical of other linguists. Although trained in the American structuralist tradition, even as a student he came to see it as too narrow as he attended lectures by Roman Jakobson, a noted European linguist and key figure in the Prague School. Some time later, he began to seriously read the works of the Prague School, which deeply impressed him. This inclusiveness became a Ferguson hallmark. He never aligned himself with any particular school of linguistics nor did he found his own. His M.A. and Ph.D. advisor, Zellig Harris (whom Ferguson greatly admired but also considered overly concerned with theory building), felt that Ferguson was much like Edward Sapir. This is evident in Ferguson's interest in various types of linguistic phenomena, particularly those considered marginal by most linguists (slips of the tongue, formulaic expressions, rhymes, proverbs, baby talk, etc.).

Rather than espouse or denounce any particular theoretical persuasion, Ferguson worked to bring diverse parties together, gently pointing out gaps in their theories and opportunities for fruitful research and cooperation. He seemed to have a gift for identifying new ground that would prove fertile. As a result, he is honored as a pioneering figure in fields such as child language, sociolinguistics, modern Arabic studies, and discourse analysis. Ferguson possessed such rare vision that even much of his early work continues to be highly relevant today.

While he ranged broadly, his work is nevertheless unified by the continuous pursuit of a better understanding of the related and fundamental topics of language universals, conventionalization, and variation. With his friend and colleague, Joseph Greenberg, he pressed for studies of a broad representative sample of various languages and language situations that would illuminate the commonalities in types of language change, structure, and use. Much of language is, however, not universal and Ferguson was fascinated with the question of how the largely arbitrary code that is the language of a given speech community comes to be shared by the individual members of the community, who in fact do not all use that language precisely alike. Some of this variation is systematic, and therefore part of the conventions of the speech community (e.g. variation that correlates with an individual's social class or gender). However, some language behavior appears to be idiosyncratic. For Ferguson, an adequate theory of human language must be capable of accounting for all of this.

Methodologically, Ferguson preferred to study a manageable linguistic phenomenon, such as politeness expressions, and consider its implications for a more general understanding of human language behavior. He found comparative case studies particularly revealing. He pushed himself and others to aim for high standards of accurate observation of actual language use, to seek for insights from different perspectives, and to include data from overlooked sources. Some of this he accomplished through his own publications—including his book reviews, some of which were quite influential and have been reprinted. In addition, however, one should not overlook his considerable influence from behind the scenes as the chair of influential committees, as a mentor, as the organizer of conferences, and as the architect of a highly influential department of linguistics.

Biography

Charles Albert Ferguson was born in Philadelphia, Pennsylvania on July 6, 1921. His early curiosity led to the study of languages and linguistics at the University of Pennsylvania, where he received his B.A. (1942) in Philosophy and his M.A. (1943) and Ph.D. (1945) in Oriental Studies; in 1947 he became a founding linguist of the Foreign Service Institute of the US Department of State and founder of the Foreign Service Arabic Field School in Beirut; in 1955–1959 he taught Arabic and linguistics at Harvard University; and in 1959 he was founding director of the Center for Applied Linguistics, Washington, DC. From 1947 to 1966 he was also actively involved at Georgetown University, and in 1967 he was founding chair of the Stanford University linguistics program. In 1971, he was president, Linguistic Society of America; in 1975 he was elected to the American Academy of Arts and Sciences; and in 1986 he was Professor Emeritus of Linguistics. He passed away on September 2, 1998.

References

Belnap, R. Kirk, and Niloofar Haeri (eds.) 1997. Structuralist studies in Arabic linguistics: Charles A. Ferguson's papers, 1954–1994. Leiden: Brill.

Dil, Anwar S. (ed.) 1971. Language structure and language use: essays by Charles A. Ferguson. Stanford, CA: Stanford University Press.

Ferguson, Charles Albert. 1959. Diglossia. Word 15. 325–40.

———. 1959. The Arabic koiné. Language 35. 616-30.

———. 1971. Absence of copula and the notion of simplicity: a study of normal speech, baby talk, foreigner talk and pidgins. Pidginization and creolization of languages, ed. by Dell Hymes, 141–50. New York: Cambridge University Press.

———. 1972. Verbs of 'being' in Bengali, with a note on Amharic. The Verb 'be' and its synonyms: philosophical and grammatical studies, Vol. 5, ed. by John W.M. Verhaar, 74–114. Dordrecht, Holland: D. Reidel.

———. 1976. The collect as a form of discourse. Language in religious practice, ed. by William J. Samarin, 101–9. Rowley, MA: Newbury House.

———. 1976. The structure and use of politeness formulas. Language in Society 5. 137–51.

———. 1995. Variation and drift: loss of agreement in Germanic. Towards a social science of language: papers in honor of William Labov, Vol. 1, Variation and change in language and society, ed. by Gregory Guy, John Baugh, Deborah Schiffrin, and Crawford Feagin, 173–98. Amsterdam and Philadelphia, PA: John Benjamins.

Ferguson, Charles Albert, and Carol B. Farwell. 1973. Words and sounds in early language acquisition: the development of initial consonantism in the acquisition of the first fifty words. Papers and Reports on Child Language Development 6. 1–60.

Ferguson, Charles A. 1998. Long-term commitment and lucky events. First person singular III: autobiographies of North American scholars in the language sciences, ed. by E.F. Konrad Koerner. Amsterdam and Philadelphia, PA: John Benjamins.

Fishman, Joshua A., Andrée Tabouret-Keller, et al. (eds.) 1986. The Fergusonian impact: in honor of Charles A. Ferguson on the occasion of his 65th birthday, 2 vols. Berlin, New York and Amsterdam: Mouton de Gruyter.

Huebner, Thom (ed.) 1996. Sociolinguistic perspectives: papers on language in society, 1959–1994/Charles A. Ferguson. New York and Oxford: Oxford University Press.

R. KIRK BELNAP

See also **Applied Linguistics: Overview; Arabic; Diglossia; Discourse Analysis; Harris, Zellig Sabbetai; Jakobson, Roman; Language Planning; Phonology; Sapir, Edward; Sociolinguistics; Structuralism**

Field Methods

Language use pervades daily life. Linguistic data are ubiquitous. Nevertheless, collection of these data requires careful consideration. Methods for gathering data are tailored to the questions being investigated. The 'classic' field situation in which a linguist starts from scratch to describe a previously unknown language is increasingly rare. More often, linguists are working within communities exploring particular elements of language use: child language acquisition, gender patterning, and linguistic variation within the community.

Data Collection

In the 'classic' field situation, linguists often work with one or two principal 'informants'. The linguist records as much as possible of naturally occurring speech, but analyzes this daily for new elements and supplements these samples with elicitations designed to test for phonemic status. Once phonemes have been designated, transcription often shifts to a practical or phonemic orthography, although details that may reflect on register, style, and/or social divisions may be lost. Classical analysis proceeds to outline the basic structures of morphology and syntax, and perhaps semantics and pragmatics. In amassing vocabulary, the collection of stories, narratives, and natural conversations may be augmented by elicitation. Word lists aid in structuring such sessions; the Swadesh list (Swadesh 1952, 1955; Hymes 1960) provides 200 core vocabulary words, chosen for universality. Forms may also be generated using canonical forms for root morphemes and systematically combining these with an inventory of the derivational affixes of the language; native speakers can then note which of these potential words are actually occurring. Linguists interested in syntactic variability may not find all the structures their models predict within their corpus of natural texts; again, elicitation may check the acceptability of sentences' potentials. Insights into the pragmatics of language use often flow from 'mistakes' in the linguist's production; laughter, awkward silences, and correction provide valuable data about appropriate forms and their felicity conditions.

Sociolinguistic fieldwork requires that the researcher record the social characteristics of the speakers. Since almost any such characteristic may impact speech, most studies restrict the variables noted to some predetermined set, often including age, educational level, gender, and socioeconomic class. Dialectology requires geographic as well as social information.

Random sampling is useful for the study of variation within a community. Even when the interview

schedule is short and minimally intrusive, cooperation is often difficult to obtain. Random means of substitution for preselected interviewees should be established.

Network interviewing typically meets with less resistance. An initial contact within the community is used to explicate the study and its purpose and to enlist further contacts. Community members may serve as brokers, interviewers, and/or researchers for the project.

Some projects may be introduced from the top down. This is especially effective in 'salvage' situations, where the community feels its language is at risk or in need of development, such as bilingual education programs. The community leaders may introduce the research and the investigator, as well as provide coresearchers, interviewers, and initial contacts.

Phone interviews are possible in many countries and can be effective, where the variable to be investigated can be quickly elicited. Snail and e-mail surveys may be used to get an 'ideal' report of people's usage, if the variable is one that is transparent and non-presupposing. Such surveys do not accurately record usage of stigmatized forms. Surveys of educators have been used in language planning in many countries; such surveys can provide data on which language of several is used by children in the classroom, by their parents, and the presence of particular forms in the linguistic repertoire, although underestimation of fluency in hegemonic languages and, paradoxically, underestimation of minority language use are common.

The 'rapid and anonymous' survey is effective for documenting linguistic variables that serve as social markers. Labov's (1996) survey of department stores in New York City used a single question about the location of a store's product to elicit the answer 'fourth floor'; the presence or absence of postvocalic /r/ in the response was correlated to socioeconomic status of the store's staff.

Linguists may analyze any corpus of language use, including broadcast and print media, electronic texts, documents, and inscribed artifacts.

Equipment

Early field linguistics was done with few recording devices beyond the pen(cil) and paper, but modern electronics allows the preservation of more complete speech events. Videotaping can capture much of the context of the speech act: age and gender of the participants, part of the physical environment, kinesics. Nonetheless, for many field situations the use of a small audio recorder is more practical, as it is often considered less intrusive. Participants tend to forget the presence of the equipment after a brief time, although the technology itself can have status implications for the speakers as well as the researcher. Recordings, digital or digitized, can be fed directly into computers, which provide spectrographic analyses, time the phonations and silences, and, in some cases, provide provisional transcription. Likewise, programs can be used to facilitate phonological analyses, finding complementary and contrastive distributions, and free variation. Text processing programs designed for interlinear translation provide consistency in glossing and ease in concordancing and dictionary creation.

Transcription Techniques

How many and which details are noted in transcription correlates with the purpose of the study. In preliminary field analysis, the researcher may be limited to transcribing isolated words. Fine phonetic details such as degree of aspiration of consonants, degree of voicing, nasalization, pitch height, loudness, length, and speed may be necessary for phonological analysis. For those languages with practical orthographies, these may be used if the problem of study is above the phonological level. Researchers involved in conversational analysis find that they need to record the speaker turns, intonation units, truncations, overlaps, pauses, and laughter. Interactional analyses often require kinesic information as well: speaker gaze, body orientation, and gestures. Some writers have used 'eye-dialect' annotations to hint at the social/regional character of the interlocutors without providing actual phonetic transcription. Because 'eye-dialect' is nonsystematic and signals 'different from standard' rather than actual production, most researchers prefer to provide sociocultural information on the speakers in introductory comments, or in comment lines throughout the texts. Such comment lines can also provide background information on the topics referenced in the recorded text. Some researchers provide key codes to selected forms that they will want to extract for analysis: e.g. prefixing negative face work with -F, noticings with N, or repairs with R. Linguistic contexts are so rich that no transcription system can capture the entire event; narrowing is a necessary adjunct of 'reduction' to writing.

Ethics

A person's identity is bound up in her speech patterns. Issues of how people speak are sensitive, especially in areas where a hegemonic language or dialect is recognized as 'standard', as such codes are regularly assumed to be correct, relegating other varieties to the status of incorrect. People who are cognizant that their

speech is being monitored may alter that speech to provide a desired self-image. This may make it difficult for the linguist to elicit or record casual or natural speech. Nonetheless, clandestine recording and use of language samples without the permission of the speakers is unethical. Where the speakers are aware of the research and its goals and support these ends, the influence of overt recording can be mitigated. Familiarity with the recorders/cameras and their increasing unobstrusiveness can yield surprisingly candid language samples. Where the language being recorded is taboo, secret, or designed for use in ritual contexts only, an understanding of how/if the data will be published must be reached. When descriptive grammars are written, care should be exercised in the example sentences used to respect the cultural norms. When possible, publications and preliminary findings should be shared with the host community. Some researchers have built in evaluation by the speakers as an additional level of data collection, bearing on the attitudinal base of language use. Language data are all around us. Linguistic field methods consist of the selection, recording, and codification of these data.

References

Briggs, Charles L. 1986. Learning how to ask: a sociolinguistic appraisal of the role of the interview in social science research. Cambridge and New York : Cambridge University Press.

Edwards Jane A., and Martin D. Lampert (eds.) 1993. Talking data: transcription and coding in discourse research. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hymes, Dell. 1960. Lexicostatistics so far. Current Anthropology 1(1).3–44.

Johnstone, Barbara. 2000. Qualitative methods in sociolinguistics. New York : Oxford University Press.

Kibrik, A.E. 1977. The methodology of field investigations in linguistics: setting up the problem. The Hague: Mouton.

Labov, William. 1966. The social stratification of English in New York City. Washington, DC: Center for Applied Linguistics.

Milroy, Lesley. 1987. Observing & analysing natural language. Oxford and New York: Basil Blackwell Ltd.

Swadesh, Morris. 1952. Lexicostatistic dating of pre-historic ethnic contacts. Proceedings of the American Philosophical Society 96. 452–63.

———. 1955. Towards greater accuracy in lexicostatistic dating. International Journal of American Linguistics 21. 121–37.

JUDITH M. MAXWELL

See also **Sociolinguistics**

Figurative Speech

Figurative speech/language is the creative manipulation of the phonological, syntactic, semantic, pragmatic structures of texts, or associations of normal language use, producing 'extra' patternings to arrive at vivid expressions and innovative ideas. It is the opposite of denotative speech/language. There are many different types of figurative speech, ranging from the repetition of a sound (alliteration) to a semantic and pragmatic contradiction in co-occurring items (oxymoron); both of these types are illustrated in the phrase 'friendly foe'.

The term 'figure' or 'figurative' comes from the Latin *figura* ('shape', 'form', or, more specifically, 'attitude', 'posture'). The concept of figures of speech is an ancient and powerful interdisciplinary one. It originated in Sicily in the fifth century $\int \chi$E as a set of conventions for speaking within a legal setting. The oratorical functions of rhetoric broadened when it was adopted in Greece. Plato attacked rhetoric; Aristotle defended it, and systematized the conventions in his *Rhetoric* (approximately 330 $\int \chi$E). The most influential Roman writer on Rhetoric was Cicero (*De Inventione,* approximately 87 $\int \chi$E). During the Renaissance, it occupied a central place in Poetics (i.e. the critical study of poetry).

The books of and on rhetoric all provide lists of terms, with definitions and examples (cf. Leech 1969; Dixon 1977). However, not all the divisions have acquired equal currency/popularity. Puttenham has classified these devices according to what sense (that of the eye, the ear, or the mind) they appeal to. A useful distinction is between 'tropes' and 'schemes'. Tropes play upon the meaning of a word/sentence. They include:

> Metaphor (referring to one thing in terms of another, e.g. He is a lion in the field. Here, the term 'lion' is used to refer to the bravery and fierceness of a human being)
>
> Metonymy (referring to something in terms of a conceptually adjacent item, e.g. *Crown* for sovereign)
>
> Synecdoche (using a part for the whole, e.g. *hand* for worker)

Simile (metaphor by explicit comparison, e.g. He is like a lion in the field)

Personification (envisaging a quality of a person, e.g. *Love* for the beloved)

Oxymoron (putting together semantically opposite terms, e.g. darkness visible)

On the other hand, schemes play upon normal sound/word order. Some of them are:

Alliteration (repetition of consonants/vowels and/or syllable within a line, e.g. the *b*reeze *b*lew)

Anaphora (repetition of the same word/phrase at the beginning of lines, e.g. the Beatitudes of the Bible that makes nine statements starting with 'Blessed are...')

Chiasmus (repetition of words or phrases in reverse order, e.g. 'If he is at Waterbath he does not care for you. If he cares for you he's not at Waterbath.')

Parison (a series of equally constructed clauses, e.g. 'It is the time of war, it is the time of care/It is the time of courage, it is the time of fear')

Zeugma (use of one word to apply to two or more words, e.g. '*Stein* her honour or her new brocade')

Another useful distinction is made linguistically, dividing the figures among phonological, orthographic, lexical, and syntactic categories. An example of a phonological figure can be alliteration, because alliteration plays with the phonemes or sound units of a language. Orthographical figures are those where the orthographical symbol, i.e. a letter/s are changed for special effects, such as 'Amerikkka' for 'America'. A metaphor is an example of a lexical figure, where a lexical item is involved in the figurative speech. A syntactic figure can be a parison, where the clauses/sentences create the figurative speech.

In addition to the examples provided above, other devices traditionally used as figures of speech include anachorism, anachronism, anticlimax, antithesis, antonomasia, apostrophe, assonance, bathos, cacophemism, climax, dysphemism, euphemism, hendiadys, hypallage/transferred epithet, hyperbaton, hyperbole, irony, litotes, meiosis/understatement, onomatopoeia, paradox, prolepsis, and pun/paronomasia (for a detailed discussion, see McArthur 1996).

It is very difficult to draw a line between figurative and literal/plain and conventional speech. The difficulty is enhanced by the fact that language itself is symbolic or figurative — we use some abstract symbols or letters to signify some real objects that we see or hear about, or some abstract ideas that we cherish.

In that sense, every word is a figurative speech. These ideas developed rapidly in the twentieth century. The classical view of figurative speech considered it to be a mere embellishment that deviates from the ordinary uses of language. This view was dominant until the end of the nineteenth century. However, the twentieth century accepted the figures to be as old as language itself.

The aim of all figurative speech is to create an impact, which may be pleasing or shocking, politically, culturally, or socially. These ends are achieved through a striking and novel imagery, through contrasts and comparisons, through juxtaposition, repetition, by evoking echoes of previous linguistic/spatial/temporal/social context(s). However, once these new concepts, associations, expressions, and comparisons are widely accepted and adopted, the figurative meaning slowly dies out or is forgotten; and the word/phrase becomes a part of the 'ordinary', 'core' language.

Figurative speech, not necessarily under that name, has remained a preoccupation in the twentieth-century literary criticism, particularly New Criticism, Stylistics, Practical Criticism, with their emphasis on micropatterns of language. Such paradigm texts as Empson's *Seven types of ambiguity* (1930) and Brooks's *The well wrought urn* (1947) are essentially reworkings of texts as structures of figurative language. Reader-oriented Criticism, like traditional rhetoric, gives focal importance to the addressee. Linguistic Stylistics, in its early days, was regarded as a modern 'descriptive rhetoric': the descriptive instruments of linguistics were deployed in the identification of rhetorical figures (Leech 1969; Fowler 1986).

There have been some plausible programs for describing and classifying figurative speech in semiotic, pragmatic, and linguistic terms (Plett 1985). The most sweeping interdisciplinary treatment of figures has been that of Jakobson (1956). He discussed metaphor and metonymy as fundamental structuring principles in such diverse fields as poetry, aphasia, and language development. Such tropes as metaphor are now widely used in analyzing how people understand and share experience (Lakoff and Johnson 1980). They have the potential to provide a promising agenda for research in Psycholinguistics and Sociolinguistics.

References

Dixon, P. 1977. Rhetoric. London: Methuen.

Fowler, R. 1986. Linguistic Criticism. Oxford: Oxford University Press.

Jakobson, R. 1956. Two aspects of language and two types of aphasic disturbance. Fundamentals of Language, ed. by R. Jakobson and M. Halle. The Hague: Mouton.

Lakoff, G., and Johnson, M. 1980. Metaphors we live by. Chicago, IL: University of Chicago Press.

Leech, G. 1969. A linguistic guide to English poetry. London: Longman.

McArthur, T. (ed.) 1996. The Oxford companion to the English language. Oxford: Oxford University Press.

Plett, H.F. 1985. Rhetoric. Discourse and literature, ed. by T.A. Van Dijk. Amsterdam: Benjamins.

Puttenham, G. 1970. The arte of English poesie, ed. by G.D. Wellock and A. Walker. Cambridge: Cambridge University Press.

DEEPSHIKHA MAHANTA

See also **anaphora; metaphor; metonymy**

Fillmore, Charles John

Charles John Fillmore studied Linguistics at the University of Minnesota, in the late 1940s, followed by a brief period of anthropology studies before ending up in the US Army. After one year of training in the United States, he spent two years in Japan, with the Army Security Agency, where he spent eight hours a day listening to short-wave broadcasts from Soviet forces in and around Vladivostok. Two years later, he was discharged from the Army and attended classes at Kyoto University while earning a living teaching English at a Buddhist girls school. When he got back home in 1956, he went to the University of Michigan in Ann Arbor, at first intending more to become a Japanologist than a linguist.

As a graduate student there, Charles J. Fillmore started out mainly as a phonologist, having received most of his phonological training just at the historical point where descriptive/structuralist phonology was giving way to generative phonology. He knew Kenneth Pike's work on Phonemics very well, as well as all of the papers on phonemic analysis that could be found in the Martin Joos Reader in Linguistics. A major turning point in his thinking was a paper by Morris Halle, a review of work by a Russian linguist, Avanesov. He had been writing a paper with the phonetician Gordon Peterson (who was his teacher then) when the Halle influence came along. The Peterson–Fillmore paper was going to be a set of principles, on the basis of which it would be possible to design an automatic program for producing a phonemic analysis of any language based on purely phonetic and distributional facts. Fillmore was not able to influence Peterson to change the direction of the paper and after that he was left on his own. But, aside from a survey paper he did with Professor Peterson and a paper with Bill Wang on 'intrinsic cues for consonant perception', and a dissertation on alternative phonemic analysis of Japanese, he never did anything else in Phonetics or Phonology.

When he got his degree, he joined William Wang at Ohio State, who was invited to start a linguistics program there. Fillmore was associated with a funded research project there called Project on Linguistic Analysis, and that is where he began doing lexical work. He taught courses in the transformationalist framework, and he became sort of evangelical about this new movement. He did some work on rule ordering (a booklet on indirect object constructions and a few related papers that nobody knows about dealt with this, and his paper on 'embedding transformations', which seems to have been the first proposal for the idea of the transformational cycle, was a part of this). The interest in verbs and their complementation patterns led up to a group of papers that formulated a proposal called 'Case Grammar', which became quickly popular. This instant popularity made him suspicious. He had received relief from teaching for ten weeks to write something on Deixis, which he somehow developed interest in while at Ohio State, but he spent the time writing 'The case for case' instead.

He was invited to spend a year at the Center for Advanced Study in the Behavioral Sciences, in Palo Alto in 1970, and at that time he wrote the 'Lectures on Deixis', which was distributed in various informal ways for 30 years and finally appeared in book form in 1997. While in California, he was invited to move to Berkeley, and he accepted. Teaching in Berkeley at first tended to deal with rule ordering arguments in Syntax, Case Grammar, and General Linguistics. In later decades, he focused on Lexical Semantics, Deixis, and Pragmatics, and all this led to a strong interest in idiomaticity and fixed expressions in general. This led in a sort of natural way to Construction Grammar, his strongest interest in the last decade of active teaching. Much of this was influenced by the interdisciplinary connections fostered by the Cognitive Science program that was just getting started on the Berkeley campus. In his thinking about Construction Grammar, he has been

influenced most by Arnold Zwicky and Paul Kay. In his lexicographic work, his strongest influence has been the British lexicographer Beryl T. Sue Atkins. And his main interest since his recent retirement from the classroom has been computational lexicography, through the direction of a project called 'FrameNet' (http://www.icsi.berkeley.edu/~framenet).

With Frame Semantics, Fillmore brings into the picture the main theoretical tenets of Cognitive Linguistics: the idea that language is an integral part of cognition, which reflects the interaction of cultural, psychological, communicative, and functional considerations.

> The basic assumption of Frame Semantics [...] is that each word evokes a particular frame and possibly profiles some element or aspect of that frame. An 'evoked' frame is the structure of knowledge required for the understanding of a given lexical or phrasal item; a 'profiled' entity is the component of a frame that integrates directly into the semantic structure of the surrounding text or sentence. (Fillmore et al. 2000:2)

The frame notion is the basis for what Fillmore (1985) calls the semantics of understanding (U-semantics), which contrasts with what may be called truth conditional semantics (T-semantics). U-semantics aims to determine what it takes for a hearer to provide an interpretation of a sentence. This is not only a compositional approach, because its operation relies on a knowledge of words, phrases, and grammatical constructions for an interpretation to be construed, but also noncompositional in that the construction process is not guided by purely symbolic operations from bottom to top. In contrast, the goal of T-semantics is to determine under what conditions a sentence may be true, where truth is determined compositionally (Fillmore 1982:230–2). In fact, for him the connection between Lexical Semantics and Construction Grammar goes beyond the matter of representation. Construction Grammar views the description of grammatical patterns and the semantic and pragmatic purposes they serve as equally important and necessary. In Construction Grammar, the semantic frame associated with a lexical item provides some of the semantic information needed for the semantic interpretation of a sentence (Fillmore 1994). Thus, lexical relations as such are not as interesting as the semantic properties that words have by virtue of their participation in frames.

His latest book, *Language form, meaning and practice* (Fillmore 2001), traces questions of language form, language use, and conventions linking form, meaning, and practice. But his latest wish is to see soon a kind of lexicographic practice that recognizes the 'frame' background of sense descriptions (cf. 'FrameNet').

Biography

Charles John Fillmore, Chuck for everyone around him, was born in St. Paul, Minnesota on August 9, 1929. He received his B.A. in Linguistics at the University of Minnesota in 1951, and his M.A. (1958) and Ph.D. (1961) at the University of Michigan. The title of his dissertation was *A system for characterizing phonological theories*. The following were his honors: Institute Professor, LSA Summer Institute, Salzburg, Austria, 1979; President, Cognitive Science Society, 1980; Election to American Academy of Arts and Sciences, 1984; President, Linguistic Society of America, 1991; Festschrifts: Masayoshi Shibatani and Sandra A. Thompson (eds.), two volumes: *Grammatical constructions: their form and meaning*, Oxford 1996; *Essays in semantics and pragmatics,* Benjamins-Amsterdam 1996; and The Berkeley Citation (UCB) 1998. He was employed from 1952 to 1955 by the US Army, Army Security Agency; from 1955 to 1957 as English teacher at Kyoto Women's School, Kyoto, Japan; from 1961 to 1970 as Assistant Professor through Full Professor by The Ohio State University; 1970–1971, Fellow, Stanford Center for Advanced Study in the Behavioral Sciences; 1971–1994, Professor of Linguistics, University of California at Berkeley; Department Chairman, Department of Linguistics, 1979–1983; Acting Director, Institute of Cognitive Studies, 1985–1989; 1994–2001, Graduate School Professor at the University of California at Berkeley; Research Associate at the International Computer Science Institute, Berkeley and director of the 'FrameNet' project. He is a member of many editorial boards.

References

Dirven, Rene, and Gunter, Radden (eds.) 1987. Fillmore's case grammar: a reader. Studies in descriptive linguistics, Vol. 16. Heidelberg: J. Groos.

Fillmore, Charles, and William S.Y. Wang. 1961. Intrinsic cues and consonant perception. Journal of Speech and Hearing Research 4(2). 130–6.

Fillmore, Charles. 1963. The position of embedding transformations in a grammar. Word 19. 208–31.

———. 1968. The case for case. Universals in linguistic theory, ed. by Emmon Bach, and Robert Harms, 1–89. Austin, TX: Holt-Rinehart-Winston.

———. 1982. Frame semantics. The Linguistic Society of Korea. Linguistics in the morning calm, 111–37. Seoul: Hanshin.

———. 1985. Frames and the semantics of understanding. Quaderni di Semantica 6(2). 222–54.

Fillmore, Charles, Paul Kay, and M. Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: the case of 'let alone'. Language 64(3). 501–38.

———. 1992. 'Corpus linguistics' vs. 'Computer-aided armchair linguistics'. Directions in corpus linguistics (Proceedings from 1991 Nobel Symposium on Corpus

Linguistics, Stockholm), 35–60. The Hague: Mouton de Gruyter.

Fillmore, Charles, and B.T. Sue Atkins. 1994. Starting where the dictionaries stop: the challenge for computational lexicography. Computational approach to the lexicon, ed. by B.T.S. Atkins, and A. Zampolli, 349–93. Oxford: Oxford University Press.

Fillmore, Charles, and Paul Kay, first author. 1999. Grammatical constructions and linguistic generalizations: the what's X doing Y? construction. Language 75(1). 1–33.

———. (with Charles Wooters, and Collin F. Baker) 2001. Building a large lexical databank which provides deep semantics. Proceedings of the Pacific Asian Conference on Language, Information and Computation, Hong Kong.

———. 2001. Language form, meaning and practice. Stanford: CSLI.

———. 1996. The pragmatics of constructions. Social interaction, social context, and language: essays in honor of Susan Ervin-Tripp, ed. by Dan I. Slobin, Julie Gerhardt, Amy Kyratzis, and Jiansheng Go. Mahwah, NJ: Lawrence Erlbaum.

———. 1997. Lectures on Deixis. Stanford, CA: Center for the Study of Language and Information, CSLI Lecture Notes, Vol. 65.

Fillmore, Charles J., et al. Construction grammar. Stanford: CSLI. To appear.

Fillmore, Charles J. 1987. A private history of the concept of case. Concepts of case, ed. by Rene Dirven and Gunter Radden. Tubingen: Gunter Narr Verlag.

Michaelis, Laura A., and Knud Lambrecht. 1996. Towards a construction-based theory of language function: the case of nominal extraposition. Language 72(2). 215–47.

Petruc, Miriam R.L. 1996. Frame semantics. Handbook of pragmatics, ed. by Jef Verschueren and Jan-Ola Östman. Amsterdam/Philadelphia: John Benjamins.

CARMEN BRETONES

See also **Deixis; Halle, Morris; Structuralism**

Finite-State Morphology

Finite-state morphology is a computational model of morphology that uses an abstract computing device called a *finite-state automaton* or a *finite-state machine*. Such an abstract machine is capable of defining a language and has a number of useful computational properties. Note that the term *language* is used here to mean a formal language, which is a set of sequences of symbols, and not necessarily a language spoken by humans. The following concise technical definition is included for completeness. Readers who do not require mathematical rigor may skip the formal definition and proceed to the intuitive explanation below.

Formally, a finite-state automaton is a quintuple $M = (\Sigma, S, s_0, F, \delta)$, where

- Σ is an *alphabet*,
- S is a finite set of *states*,
- s_0 is a designated *start state* within S,
- F is a subset of S called the set of *final states*, and
- δ is the *transition function* $\delta : S \times \Sigma \rightarrow S$.

Such an abstract machine is capable of defining a language through an extension of the transition function, $\delta^* : S \times \Sigma^* \rightarrow S$, that extends δ to sequences of members of Σ^*, the set of sequences of symbols taken from Σ. We define δ^* as follows:

- Let $\sigma_1 \ldots \sigma_n \in \Sigma^*$ be any sequence in Σ^*.

- Let λ denote the empty sequence consisting of no symbols. Then
  - $\delta^*(s, \lambda) = s$ and
  - $\delta^*(s, \sigma_1 \ldots \sigma_n) = \delta(\delta^*(\sigma_2 \ldots \sigma_n), \sigma_1)$ when $n > 0$.

The language $L(M)$ defined by M is the set $L(M) = \{\sigma \in \Sigma^* \mid \delta^*(s_0, \sigma) \in F\}$.

Intuitively, one can think of a finite-state automaton as a diagram where the set of states are circles and the transition function is a set of arrows connecting states as in Figure 1.

Each state has one arrow pointing out of it for each element of the alphabet. States are generally labeled for identification, and the start state and the set of final states are marked in some way. Such a diagram can be thought of as a machine by imagining a bug sitting on the start state. As the symbols of a particular sequence, σ, are called out, the bug crawls along the arrow labeled with that symbol to the next state. If, after the entire sequence has be called, the bug is sitting on a state that is a final state, then the sequence σ is in the language defined by the automaton. In the example shown in Figure 1, the language is the set of sequences of a's and b's that contain an odd number of a's. A finite state machine used in this way is called a *finite-state recognizer* since it recognizes sequences of a particular language (see Figure 1).
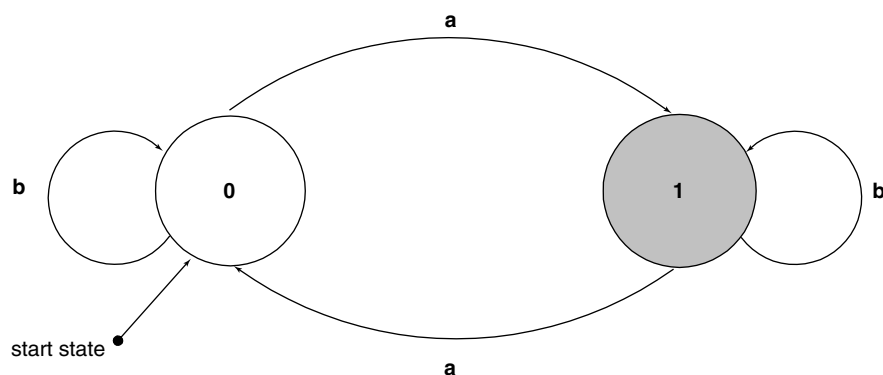
Figure 1. Odd number of a's (final state is shaded).

Although every state must have an arrow for each symbol of the alphabet, we generally draw the diagram with an assumed sink state that is not drawn. The sink state is one where all arrows simply point back to the same state. When an arrow is missing in the diagram, it is assumed to point to a sink state that is not in the set of final states.

If, on every arrow, we put a pair of symbols, we have what is called a finite-state transducer. Intuitively, we think of the first symbol as the input symbol and the second symbol as the output. This provides us with a kind of translator. It is finite-state transducers that are generally used for finite-state morphology.

Finite-state machines are important because they are computationally simple to implement and have a number of convenient computational properties. For example, the amount of memory required to determine if a given sequence is in the defined language is independent of the length of the sequence. Another useful property is that both finite-state recognizers and finite-state transducers can be easily combined to form the union or intersection of languages. Hence, specific morphological rules can be modeled independently by separate finite-state transducers and then combined into a single machine. Although not all languages can be defined by finite-state machines, the morphology of many human languages can be modeled quite successfully with finite-state models.

In order to use finite-state transducers for morphology, the alphabet is defined as the set of morphemes and the diagram defines the sequences of morphemes that properly define a word for a given human language. The transducer defines a 'translation' between the morphological rules that are used to construct the word and the surface realization of the word.

By far the most prevalent model for finite-state morphology is two-level morphology as proposed by Kimmo Koskenniemi, who used the model originally to describe Finnish morphology. Koskenniemi uses a set of finite-state transducers to produce a mapping between the surface graphical representation of a word and its phonological/morphological representations. In Koskenniemi's model, the multiple machines are not usually combined into a single machine, but operate in parallel, with each machine representing one morphological rule. For example, using a variant of Koskenniemi's original notation, the rule

y/i $\Leftrightarrow$ $\lozenge$[bcdfghjklmnpqrstvwxyz]~ • [e][s]

would define a rule that requires a 'y' that is preceded by a consonant and followed by the plural ending 's' being replaced by 'i' in the surface representation. The rule is essentially the rule learned in grammar school: 'To form the plural of a word that ends in a 'y' preceded by a consonant, then change the 'y' to 'i' and add es'. In the notation here, the symbol '~' denotes an instance of the left-hand side of the rule and the symbol • denotes a morphological boundary. The square brackets enclose alternative symbols so that each set of square brackets represents one symbol from among the alternatives.

Koskenniemi originally used his model for Finnish and hand translated each rule into a finite-state transducer. Later programs carry out the translation by computers. The two-level model is very powerful and can handle the morphologies of at least a large portion of the world's human languages.

References

Hopcroft, Jjohn E., and Jeffrey D. Ullman. 1979. Introduction to automata theory, languages, and computation. Reading, MA: Addison-Wesley Publishing Company.

Kornai, András (ed.) 1999. Extended finite state models of language. Cambridge: Cambridge University Press.

Roche, Emmanuel, and Yves Schabes (eds.) 1997. Finite-state language processing. Cambridge, MA: MIT Press.

Sproat, Richard. 1992. Morphology and computation. Cambridge, MA: MIT Press.

ALTON SANDERS

See also **Morphology**

Finnish and Finnic Languages

Finnic (also Fennic, Balto-Finnic) belongs to the *Finno-Ugrian languages* (also Finno-Ugric), one of the two branches of the *Uralic* languages (the other consists of Samoyed languages). In addition to Basque and the Indo-European languages, it is the only other language family represented in Europe. The concept Finno-Ugrian is based on the two dominant Finno-Ugrian languages: Finnish (Finnic) and Hungarian (Ugric). There are two theories about the original Finno-Ugrian homeland. It was situated on the western side of the Uralic mountains, at River Volga's great turn south, or is believed to have consisted of a continuum of Finno-Ugrian populations from the Uralic mountains to the Baltic sea. The latter view, which has gained increasing support, partly stems from a reevaluation of the history and development of the Finnic languages.

The Finnic languages are spoken around the Gulf of Finland. Finnish is spoken north of the Gulf, in Finland; Estonian is spoken south of it; and east of the Gulf, several minor Finnic languages are spoken: Votic (also Votian, Vote), Ingrian (also Izhorian), Karelian, Ludian, and Veps (also Vepsian). In northern Latvia, a few people still speak Livonian. Ingrian, Votic, and Livonian are likely to lack mother-tongue speakers in approximately one generation.

There have been continuous contacts in the Baltic area among early Indo-European, Baltic, Germanic, and Finno-Ugrian people since about 500–2000 BCE. Estonia has probably been inhabited by Finno-Ugrian people since 2000 BCE. Three major ancestral (proto) forms of Finnic were probably established around 300–1000 BCE. The area north of present-day St. Petersburg, the Karelian Isthmus, the area north of Lake Ladoga, and that between Ladoga and Lake Onega have housed people who later became the Karelians, the Ingrians, and possibly the Veps. By 800 CE, the Finnic languages had reached the geographical distribution that became the basis of their autonomous linguistic developments.

The early phases of the Finnic languages are not well known. The people have been known since the first century CE. The same name has, however, been used to refer to different people. The use of Fenni (by Tacitus in his *Germania*, 98 CE) was apparently used for the Saami and by others, later, for the Finns. The use of Aestii probably referred to Baltic people initially and only later became reserved for Estonians. The Saami languages are related to, but not counted as, Finnic languages. The Saami presumably shifted their language to or were extensively influenced by Proto-Finnic. The genetic origins of the Saami remain a riddle. With the expansion of the Finns in Finland, the Saami were pushed northward.

The Finnic influences have followed two main lines. There have been west–east contacts between Finnish and Karelian/Ingrian (less with Votic and Veps) and north–south contacts between Finnish and Estonian. Estonian, in turn, has been in intensive contact with Votic and Livonian. Between the periphery languages, for example, Finnish and Veps or Livonian, there are considerable linguistic differences and restricted mutual intelligibility.

The external linguistic contacts have had four main historical directions: Swedish influenced Finnish, German influenced Estonian, Russian influenced Karelian (as well as Ludian, Ingrian, Votic, and Vepsian), and Latvian influenced Livonian. Reversed local influences have also occurred, but to a lesser extent.

The Finns and Estonians were Christianized by Germanic people in the twelfth and thirteenth centuries. They remained Catholics until the Reformation. The Karelians, Votes, and Ingrians were Christianized from the south and have remained Eastern Orthodox. The Votes and Ingrians, as well as the Veps, have belonged to the political spheres of Novgorod, Moscow, and, later, Russia. A peace treaty in Pähkinäsaari in 1323 created the first national border of Finland. This border between Sweden and Finland and Novgorod and Russia has been modified many times, especially in 1617, when the Swedish conquest of the Baltic-Finnic area in the early seventeenth century caused an exodus of Orthodox Finnic people into Russia. The border has developed into a language boundary between Finns and speakers of Karelian, Ingrian, and Votic.

Finnish is spoken by approximately 5.1 million speakers in the Baltic area, including 250,000 speakers in Sweden, 10,000 in northern Norway, and 60,000 in Russia. Approximately 94% of Finland's population (5.1 million) are Finnish-speakers. Since 1919, Finnish and Swedish have been the national official languages of the Republic of Finland. Since 2000, Finnish has been an official minority language in Sweden. From time to time, Finnish has had official status in Karelia. Migrant communities can be found in the United States and Canada (100,000–150,000 speakers), Germany

(30,000), Estonia (13,000), Norway, Denmark, Australia, and elsewhere.

Finnish was codified in the sixteenth century and recodified during the nineteenth century. Up to 1540, it was called *old Finnish*. The period of *old written Finnish* continued into the 1820s. The period between 1820 and 1870 is called *early modern Finnish*. During that period, which coincided with that of the Grand Duchy of Finland (1809–1917) under the Russian Empire, a romantic, nationalistically based discussion on the origins, the future direction, and status of Finnish was largely settled. The birth of *modern Finnish* is set to around 1870, when Aleksis Kivi's novel *Seven brothers* was published. During the nineteenth century, the older, analytic Swedish-influenced language changed into the modern form of Finnish. The dialects were used as a source of relexification (renewal of the vocabulary) and grammar development.

Early Finnish writing was based on southwestern (Turku) dialects. The New Testament (1547) was translated by the cleric Michael Agricola, who lived in Turku, the political and religious center of Finland (Stockholm was the capital). The southwestern impact was weakened in the translation of the Bible (1642). Of great importance was the collection and compilation of the Kalevala epic (by Elias Lönnrot; it was completed in 1849), which had eastern origins (Savo and Karelian). The written standard formed a compromise based on older western and later eastern sources. A new spoken standard was also developed. The orthography closely matched the spoken standard: one sound largely corresponded to one sign/letter, and one letter corresponded to one way of pronouncing it. This ideal relationship does not persist today.

Finno-Ugrian studies have contributed important hypotheses about the older cultural development in the area. The contacts with Indo-European languages were established early, maybe around 2000 BCE. Among the oldest Indo-European loanwords in Finnish, one finds *nimi* 'name', *vesi* 'water', and *kesä* 'summer'. Some of the oldest Indo-European vocabulary belongs to the core vocabulary. There has also been a period of close contacts with Aryan/Iranian languages, the time of which is uncertain: *jumala* 'God', *syntyä* 'be born', and *taivas* 'heaven' stem from these contacts. Among the old loanwords are also Baltic ones: *hirvi* 'elk', *lohi* 'salmon', *heinä* 'hay', *kirves* 'ax', *seinä* 'wall', *silta* 'bridge', and *heimo* 'tribe'. The Baltic influence has been interpreted as a result of intimate contacts and interethnic marriages. Another long-term influence was Germanic: *juusto* 'cheese', *leipä* 'bread', *ranta* 'shore', *kulta* 'gold', *rauta* 'iron', *kauppa* 'shop', *raha* 'money', and *kuningas* 'king'. Slavonic influence is much more recent (600–800 CE): *lusikka* 'spoon', *ikkuna* 'win-

dow', *pappi* 'priest', *pakana* 'pagan', and *risti* 'cross'. From the twelfth century, Scandinavian/Swedish influence has been quite strong on Finnish: *katu* 'street', *laki* 'law', *helvetti* 'hell', *lasi* 'glass', and *tuoli* 'chair'. Swedish has been a mediator of western European cultural influences to Finnish throughout their common history.

Some of the older core vocabulary in Finnish, such as *kieli* 'tongue; language', *nuoli* 'arrow', *veri* 'blood', *käly* 'sister-in-law', *silmä* 'eye', *suksi* 'ski', *kala* 'fish', *päivä* 'day', *koivu* 'birch', *kaksi* 'two', *viisi* 'five', *elää* 'live', and *kuolla* 'die', probably stem from the Uralic protolanguage period (approximately 4000–5000 BCE). Several of the words reflect typical features of Finno-Ugrian languages.

Some Finnic and Finno-Ugrian languages have *vowel harmony*, for example, Finnish (Estonian lost it around the seventeenth century). Vowel harmony restricts the combination of vowels in a word and its endings according to the principle of assimilation. Front vowels (y, ö, ä) combine with each other, and back vowels (u, o, a) combine with each other. For example, *käly* 'sister-in-law' is based on front vowels, and *kuolla* 'die' is based on back vowels. The vowels /i/ and /e/ are considered neutral.

In Finnish and the Finno-Ugrian languages, words can be attached to each other in long rows (agglutination) to express the same content and grammatical relations that languages such as English express with separate words. The Finnish two-word phrase *Menisimmekö iltakävelyllemme?* corresponds to the English sentence 'shall we go for our evening walk?'

Finnish and Finnic languages lack definite and indefinite articles. The phrase *pimeässä metsässä* 'in + dark + forest' may mean 'in *a* dark forest' and also 'in *the* dark forest'. Finnish has taken steps on the way to develop a definite article during the last two centuries, *se* 'it', and an indefinite article, *yksi* 'one'.

Finnic languages lack grammatical gender. The third-person singular pronoun *hän* in Finnish is identical for 'he' and 'she'.

Finnish has two main dialect groups, the western and eastern dialects, which can be subdivided: Häme dialects, heirs of early Proto-Finnish; southwestern dialects with clear traces of Scandinavian and Estonian; Ostrobothnian dialects; Savo dialects, probably developed from Karelian; north Finnish dialects developed from trade and contacts with the north from Karelia, Häme, southwestern Finland, and, later, Savo areas; and southeastern dialects. Meänkieli, the other official minority language in northern Sweden, originates from the 'mixed' northern dialects of Finland. The Finnish dialects in Tornedalen in turn became the main source for Kven Finnish in northern Norway (seventeenth century). For Finnish dialects, few

problems of mutual intelligibility exist, except possibly between the southwestern and the eastern Savo dialects.

Capital Helsinki speech plays a central role in present-day Finnish. Helsinki has experienced several language shift periods between Swedish and Finnish since the sixteenth century. It has received many waves of speakers of different dialect backgrounds. Western dialects, especially the southern Häme dialects, have dominated. At the end of the 1800s, the influx of both Finnish-speaking and Swedish-speaking labor forces created a basis for extensive language contacts. Finally, standard spoken Finnish, which was established in the young often bilingual bourgeoisie of the late 1800s and early 1900s, has contributed to Helsinki speech. These different linguistic origins have created a mix not found elsewhere. For approximately 40–50 years, Helsinki speech has been an overt prestige model for the rest of Finland.

Estonian has two main dialects: the northern dialects (with Tallinn as the main city) and the southern dialects (main city Tartu), including the southeastern Võru dialect. The orthography is based on the northern dialects, but the southern dialects have had a writing of their own. The main dialect division may reflect the two early Finnic protolanguages south and east of the Gulf. Estonian is the third largest Finno-Ugrian language, with approximately 1 million speakers out of the 1.5 million inhabitants in Estonia. The remainder (0.4 million) mainly speaks Russian. Estonian was formally an official language, together with Russian, during the Soviet rule, but this changed after independence in 1991. In 1995, the new Language Law made Estonian the sole official language of the Republic of Estonia. Estonian communities live in northwestern Russia, Finland (10,000), and Sweden (10,000). Migrant communities live in North America and Western Europe. As a result of Russian and Soviet time deportations, Estonians also live in Siberia and the Far East.

The translation of the Catechism in 1535 by Johann Koell (in Wittenberg, Germany) introduced written Estonian. During the following centuries, Estonian books were mainly written by German speakers; hence, German had an extensive influence on written Estonian. Beginning in the nineteenth century, German influence faced a nationalistic 'purification' similar to that of Swedish in Finland. Estonians were inspired by the Finnish attempts, to the extent that a folk epic similar to Kalevala was compiled: Kalevpoeg (in 1862; by Friedrich Reinhold Kreuzwald). The 'nationalization' of Estonian brought about greater linguistic distance to Finnish, because the language activists in Estonia and Finland preferred different solutions when they created new words for Germanic loanwords. The lexical influence of Russian grew during the twentieth century, but its linguistic impact remains smaller than that of German.

Linguistic differences between Estonian and Finnish include the central high Estonian vowel /õ/, võõras, cf. Finnish vieras, 'stranger; guest'. Estonian also lost word-final vowels: Finnish silmä vs. Estonian silm 'eye', Finnish piimä 'sour milk' vs. Estonian piim 'milk', Finnish Raamattu 'Bible' vs. Estonian raamat 'book'. The last two pairs also exemplify the many shifts in meaning. Estonian lacks the Finnish diphthongs, for example: Finnish (minä) syön corresponds to Estonian ma söön, 'I eat'.

Estonian has resisted the dominance of both German and Russian. The great social distance between the German speakers and Estonians may have been one reason, because Estonians experienced a harsh serfdom under Balto-German landlords up to the nineteenth century. Another reason may be that Peter the Great created an administrative and cultural center east of Estonia when he founded St. Petersburg, his window to the west at the Baltic Sea. Other Finnic languages in that area were submersed by Russian dominance. For example, Votic and Ingrian today rely on some dozens of speakers. The area around St. Petersburg has also been a battlefield during several wars between Sweden/Finland and Russia and between Russia and Germany. Genocide of Finnic people (especially Ingrians) during some of the Soviet regimes and Russification of all citizens of the Soviet Union have contributed to the language death of the Finnic neighbors of Estonian.

Karelian is the third largest of the Finnic languages. It is believed to have developed more independently from the ninth to eleventh century CE. Proto-Karelian made up the basis for Ingrian, but Karelian itself has developed in close contact with Veps. Karelian consists of four major dialects, which are partly an effect of different language contacts. North Karelian dialects are close to the eastern Finnish dialects. South Karelian is spoken west and north of Lake Ladoga. Aunus is spoken southeast of south Karelian, in and around the town of Aunus. Ludian is spoken west of Petrozavodsk. Ludian is sometimes described as a contact language of Karelian and Veps. As a result of the exodus into Russia in the seventeenth century, some Karelian 'islands' developed close to Kalinin. These so-called Tver-Karelians still speak Karelian.

During the 1930s, three attempts were made to create a standard Karelian orthography. Karelian is a minority language within its own autonomous Karelian Republic (approximately 10% of the 800,000 inhabitants speak Karelian). Their distribution is uneven, however: Karelian has been better maintained in rural areas. The dialect split has made the choice of

Numerals 1–10 in Finnish, Estonian, Veps, Northern Saami, and Hungarian

	Finnish	Estonian	Veps	Northern Saami	Hungarian
1	yksi	üks	üks′	okta	egy
2	kaksi	kaks	kaks′	guokte	kettő
3	kolme	kolm	koume	golbma	három
4	neljä	neli	nel'l'	njeallje	négy
5	viisi	viis	viž	vihtta	öt
6	kuusi	kuus	kuz′	guhtta	hat
7	seitsemän	seitse	siičmen	čiežá	hét
8	kahdeksan	kaheksa	kahcan	gávcci	nyolc
9	yhdeksän	üheksa	ühcan	ovcci	kilenc
10	kymmenen	kümme	kümn′en	logi	tíz

one variety and one writing system difficult. Speakers of the northern dialects and the notable group of Finnish speakers in Karelia have preferred standard Finnish orthography. The fact that all Karelians know Russian, and the strong Orthodox faith, are other factors complicating the introduction of a Latinized Karelian writing.

Veps is often described as an archaic Finnic language. In historical time, it remained in contact with the eastern Finno-Ugrian languages. It is spoken south of Ludian and west and south of Lake Onega and Lake Valgjärvi. The number of speakers of Veps is estimated at 10,000 to 12,000. Three of four main dialects have survived: North, Central, and South Veps. The number of speakers of Veps and its area have diminished continuously. The post-Soviet society has made it possible to support and study Veps (also for western Finno-Ugrists). In-depth studies of Veps would cast additional light on the contacts among the Indo-European, Baltic, Finnic, and other Finno-Ugrian people. They could also add new aspects to the study of the potential development of a Baltic Sprachbund. The study of and support to Veps are urgent tasks because of its ongoing language shift.

References

Abondolo, Daniel (ed.) 1998. The Uralic languages. London: Routledge.

Anttila, Raimo. 1988. Historical and comparative linguistics. Amsterdam: John Benjamins.

Fogelberg, Paul (ed.) 1999. Pohjan poluilla. Suomalaisten juuret nykytutkimuksen mukaan (Tracing the North. The roots of the Finns in the light of recent research), Bidrag till kännedom om Finlands natur och folk, 153. Helsinki: Finska Vetenskapsocieteten.

Häkkinen, Kaisa. 1996. Suomalaisten esihistoria kielitieteen valossa (The pre-history of the Finns as described by linguistics), Tietolipas 147. Helsinki: Suomalaisen Kirjallisuuden Seura.

Heinsoo, Heinike. 2000. On the formation of the Finnic linguistic situation. The roots of peoples and languages of Northern Eurasia, vols. II and III, ed. by Ago Künnap, Fenno-Ugristica 23. Tartu: Division of Uralic Languages, University of Tartu.

Karlsson, Fred. 1983. Finnish grammar. Porvoo, Helsinki and Juva: WSOY.

Korhonen, Mikko. 1981. Finno-Ugrian language studies in Finland 1828–1918. Helsinki: Societas Scientiarum Fennica.

Laakso, Johanna (ed.) 1991. Uralilaiset kansat. Tietoa suomen sukukielistä ja niiden puhujista (The Uralic people. Facts about the language relatives of Finnish and their speakers). Porvoo, Helsinki and Juva: WSOY.

———. 1999. Karhunkieli. Pyyhkäisyjä suomalais-ugrilaisten kielten tutkimukseen (The language of the bear. Strokes of Finno-Ugrian language studies), SKST 729. Helsinki: Suomalaisen Kirjallisuuden Seura.

Laanest, Arvo. 1982. Einführung in die ostseefinnischen Sprachen. Hamburg: Buske.

Lallukka, Seppo. 1990. The East Finnic minorities in the Soviet Union. Suomalaisen tiedeakatemian toimituksia 252. Helsinki: Suomalainen tideakatemia.

Lehiste, Ilse. 1988. Lectures on language contact. Cambridge, MA and London, England: MIT Press.

Niemi, Jussi, Terence Odlin, and Janne Heikkinen, (eds.) 1998. Language contact, variation and change. Studies in language, Vol. 32. Joensuu: Faculty of Humanities, University of Joensuu.

Pyöli, Raija. 1996. Venäläistyvä aunuksenkarjala. Kielenulkoiset ja –sisäiset indikaattorit kielenvaihtotilanteessa (The Russification of Aunuksenkarjala. Language external and internal indicators in a language shift situation), University of Joensuu Publications in the Humanities, Vol. 18. Joensuu: University of Joensuu.

Sinor, Denis (ed.) 1988. The Uralic languages: description, history and foreign influences. Handbook of Uralic studies, Vol. 1. Leiden: Brill.

Stipa, Günter J. 1990. Finnisch-ugrische Sprachforschung: Von der Renaissance bis zum Neopositivismus, Mémoires de la Société Finno-Ougrienne, Vol. 206. Helsinki: Suomalais-ugrilainen seura.

Uibopuu, Valev. 1988. Finnougrierna och deras språk: kapitel om de finskugriska folkens förflutna och nutid (The Finno-Ugrians and their languages: chapters about the past and present of the Finno-Ugrian people). Lund: Studentlitteratur.

Virtaranta, Pertti, and Seppo Suhonen. 1978. Lähisukukielet: Näytteitä uralilaisista kielistä, III, (The closely related languages: examples of the Uralic languages, Vol. III) Suomi 119.3. Helsinki: Suomalaisen Kirjallisuuden Seura [Texts and cassettes of the Finnic languages].

JARMO LAINIO

See also **Case; Hungarian and Ugric Languages; Language: Contact - Overview; Vowel Harmony**

Firth, John Rupert

The British linguist John Rupert Firth (1890–1960) has traditionally been associated with two quite different approaches in linguistics. The first approach relates to a sociological linguistic theory called the *Theory of the context of situation*, influenced by the works of the anthropologist Bronislaw Malinowski. The second is a technique of phonological analysis developed by the so-called London School. However, a detailed study of Firth's work and teaching (mainly from 1930 to 1957) forces us to describe his influence in linguistics in somewhat different terms. Throughout his works, Firth developed a linguistic theory of his own that places special theoretical emphasis on two levels of analysis: the social level (equivalent to the *pragmatic* level, in modern terms) and the phonological one.

Firth explained his theory on language in *Speech* (1930) and *The tongues of men* (1937). Both works present functionalist and relativist thinking in the study of language, which Firth called 'descriptive linguistics'. Firth developed a theory for language in action, not a universal theory of human linguistic ability. These books are a presentation of his thinking similar to Sapir's *Language* (1921) and Bloomfield's *Language* (1933), but Firth did not expound a method of linguistic analysis. Rather, he wrote for the general public and showed a multidisciplinary theoretical approach, which he maintained throughout his works. Thus, these books develop a study of language from the biological, historical, sociological, and linguistic point of view. The rest of Firth's work—its greatest part—is published as highly specialized theoretical articles or collections of articles. Firth's style has been described as obscure, mainly because he rarely focused on one topic throughout any given article. His essays are usually an amalgam of theoretical knowledge, readings, and opinions.

Firth understood language as a system of signs that we use to do things; that is to say, language works in different ways in different contexts. For Firth, language is somewhat magical, and this linguistic power is what we learn from the day we are born. All the events in a speaker's life (society, culture, roles, personalities, etc.) dictate the form of this person's language. Some critics detect here a behaviorist attitude in Firth's writing, inherited from Malinowski's works and teaching. However, one could point out that this may be just one of the many functionalist aspects of his thinking, and that there is a relationship between Malinowski (and later Firth) and the American prag-matism in the works of G.H. Mead, J. Dewey, and G. de Laguna.

Descriptive linguistics identifies the social component of language as a critical aspect of language, and Firth introduced a linguistic category encompassing the elements that usually occur in a communicative situation. He labeled this category *Context of situation*, and its elements are:

(1) The participants: persons, personalities, and relevant features of these.

    (a) The verbal action of the participants.
    (b) The nonverbal action of the participants.

(2) The relevant objects and nonverbal and nonpersonal events.
(3) The effect of the verbal action.

Firth used a recurring metaphor—the dispersion of light into a spectrum—to explain another important part of his theory: levels of analysis. He described linguistics as the prism through which light (the language) passes. The result of this division—or dispersion—are the different modes of meaning that we can find in the linguistic analysis of an utterance. These modes of meaning must be described by their own linguistic categories, which establish different levels of analysis (e.g. phonetics vs. phonology vs. morphology, etc.). The different meanings must be revealed and analyzed by the linguist, keeping in mind a very important premise in firthian theory: that Meaning is Function (cf. Wittgenstein's *Philosophical investigations*, 1954).

Thus, descriptive linguistics studies different meanings on the different levels of analysis, but not necessarily in a hierarchical way (from phonetics to semantics or vice versa). Apart from the context of situation level, Firth posits the lexical, grammatical (syntactical and morphological level), phonological, and phonetic levels, but he is committed to a specific number of levels.

Two theoretical concepts developed by Firth now form part of the technical vocabulary of linguistics: collocation and colligation. The first is used at the lexical level and refers to a word that acquires part of its meaning by the company of the words it usually keeps. The second term is colligation, very similar to agreement, which is used at the grammatical level to refer to the syntactical relationship that an element maintains within a construction with others.

The theoretical points of view that Firth maintained in his works are very peculiar. Since the publication of De Saussure's *Cours de linguistique générale* (1916), language has been divided into systems and structures with their respective paradigmatic and syntagmatic relationships. One concept, however, separates Firth (and later the London School) from the rest of the linguistic paradigms: polisystemicity. Polisystemic means 'several systems', and this concept is applied to different aspects of linguistic analysis. For example, a language can contain several phonetic systems: a native system with its own phonetic patterns, and systems of foreign phonetic patterns borrowed from other languages. At the same time, a language can contain several subsystems within its structures: a consonant–vowel–consonant pattern (CVC) is a structure that contains a system for initial consonants C-, a system for medial vowels -V-, and a system for final consonants -C. This polisystemic approach is opposed to the monosystemic approach (typical of structuralism), where only one system is posited for the whole language (a good example is the hypothesis of a 'phonological system' of a language containing a closed set of phonemes).

Firth criticized this monosystemic and paradigmatic approach because it provides only an incomplete analysis of a language. His polisystemic approach, in contrast, studies not only the paradigmatic relationships but also the syntagmatic ones. These are the theoretical underpinnings of the main categories that Firth established at the phonological level: phonematic unit and prosody. These two categories form the conceptual basis of Firth's phonology, and later of the Prosodic Analysis developed by the London School.

Phonematic unit is a phonological category, applied at the phonological level. Sound structures are represented using the phonematic units 'C' for consonants and 'V' for vowels. Each one of these units is, at the same time, the representation of a system given at that place in the structure where some potential elements appear. Phonematic unit is thus the expression of a paradigmatic or vertical relationship. On the other hand, prosody, far from classical definitions (related to poetry, metrics, or suprasegmentals), is a syntagmatic or horizontal relationship given in a structure, i.e. something that happens at the same time in the structure as a whole and characterizes it. In fact, a prosody is usually the guide for setting up a structure. These concepts are implicit in Firth's early works but are most clearly developed in his 1948 article 'Sounds and prosodies'. This article is known as the foundation of the *Prosodic analysis* (later *Firthian prosodic analysis*, FPA), a technique of phonological analysis almost solely used by the London School. The most important work of this school was done during the 1950s and 1960s by authors such as E. Henderson, J. Carnochan, K. Sprigg, F.R. Palmer, and R.H. Robins. Although some bibliographies claim that FPA failed, it is still being developed (with some theoretical changes due to recent phonological studies) in some parts of the United Kingdom.

Thus, Firth developed a linguistic theory that found different applications on different levels of analysis. Firth's influence, although important, is not well known. While his name is mainly associated with the London School and *Firthian prosodic analysis, systemic-functional grammar*, which is associated with the work of M.A.K. Halliday, he is also known as *Neo-Firthian*. More complete works on Firth and FPA are currently becoming available, which, in some sense, is a vindication of his theoretical contribution to linguistics.

Biography

John Rupert Firth was born in Leeds, England, Great Britain, on June 17, 1890. He did his B.A. (1911) and M.A. (1913), both in history. He was a lecturer in History at the City of Leeds Training College in 1913, and that year joined the Indian Education Service. He was Professor of English at the University of the Punjab, Lahore (now Pakistan) (1920–1928), and Senior Lecturer in the Department of Phonetics at University College, London (1928–1938), holding meanwhile a number of part-time appointments. He had an Assistantship in Sociology of Languages at the London School of Economics (1928), where he worked with Bronislaw Malinowsky, was Senior Lecturer in the School of Oriental and African Studies in 1938, and Reader in Linguistics and Indian Phonetics in 1940.

Firth held the first chair of General Linguistics in Great Britain (1944) at the University College, London. He was also Head of the Department of Phonetics and Linguistics since 1941. Although he retired in 1955, he continued attending the Philological Society meetings, of which he was a member since 1933, President from 1954 to 1957, and Vice-President in 1959. He died on December 14, 1960.

References

Anderson, S.R. 1985. Phonology in the twentieth century. Chicago: University of Chicago Press.

Bazell, C.E., J.C. Catford, M.A.K. Halliday, and R.H. Robins. (eds.) 1966. In memory of J.R. Firth. London: Longmans.

De Beaugrande, R. 1991. Linguistic theory. The discourse of fundamental works. London: Longmans.

Firth, John Rupert. 1948a. Sounds and prosodies. Transactions of the Philological Society. 127–52.

———. 1950. Personality and language in society. Sociological Review 42(2). 37–52.

———. 1957. Ethnographic analysis and language with reference to Malinowski's views. Man and culture: an evaluation of the work of Bronislaw Malinowski, ed. by Raymond Firth, 93–118. London: Routledge & Kegan Paul.

———. 1957. Papers in linguistics, 1934–1951. London: Oxford University Press.

———. 1957. A synopsis of linguistic theory, 1930–1955. Studies in linguistic analysis, Special volume of the Philological Society, 1–32. Oxford: Basil Blackwell.

———. 1964. The tongues of men and speech. London: Oxford University Press, Reprinted 1966 and 1970.

Harris, Roy (ed.) 1988. Linguistic thought in England 1914–1945. London: Duckworth.

Kelly, J., and J.K. Local. 1989. Doing phonology. Manchester: Manchester University Press.

Langendoen, D.T. 1968. The London School of Linguistics. A study of the linguistic views of B. Malinowski and J.R. Firth. Cambridge, MA: MIT Press.

Palmer, F.R. (ed.) 1968. Selected papers of J.R. Firth (1952–1959). London: Longmans.

———. 1970. Prosodic analysis. London: Longmans.

Robins, R.H. 1963. General linguistics in Great Britain 1930–1960. Trends in modern linguistics, ed. by Mohrmann, Norman and Sommerfelt. Antwerp: Spectrum.

ELENA BATTANER-MORO

See also **Phonology; Pragmatics; Prosody**

Fishman, Joshua A.

With his prolific research work spanning almost four decades, Joshua Fishman has been recognized as a leading sociolinguist/sociologist of language. He has been influential in scholarly studies of the relationship between language and society—at the macro level (i.e. the broad societal level) as against micro (i.e. the individual level). His major theoretical contribution is the concept of 'domains' (i.e. a larger unit than situation, e.g. that of family, work) of language use and the extension of the concept of 'diglossia' (i.e. the existence of two or more varieties of the same dialect/language) to include bilingualism. The fact that three volumes of scholarly articles were brought out on his 65th birthday indicates the extent of his academic influence.

Fishman's first scholarly article appeared in *Yidishe Shprakh* in 1947. It was an outcome of his ardent interest in the study of his mother tongue, Yiddish. In 1949, he received a prize from the Yiddish Scientific Institute for an unpublished monograph on bilingualism. An extended study of minority languages and bilingualism started in response to a routine questionnaire circulated by the US Census Bureau for the national census of 1960, for which he suggested a revision of the language questions. This initiated his campaign to establish the importance of collecting data on the status of non-English languages in the United States. Eventually, it led to the publication of the book *Language loyalty in the United States* (1966), the major reference work in the preparation of the Bilingual Education Act (Senate Bill 428). It also established a new area of research in Sociolinguistics: language maintenance and shift as a field of study.

Broad perspectives on the issues involved in language planning typify Fishman's work—he has probed into interrelated topics such as the relationship of language, ethnicity, nationality, and state (see Fishman 1972); language maintenance and shift (1964); bilingualism and diglossia (1967); languages of wider communication (1989); the stages involved in adequate language planning (1974); and issues of corpus planning such as the development of vocabulary in Israel (Fellman and Fishman 1977). His edited book, *Handbook of language and ethnic identity*, is a collection of 28 unpublished papers that provide a kaleidoscopic vision of a variety of social issues—including Economics, History, Political Science, Ethnography, Sociology, Sociolinguistics, Psychology, and Religion.

A macrosociological and quantitative (i.e. statistical data collection) approach is the distinguishing feature of Fishman's work. His work is characterized by meticulous analysis of large bodies of data collected in major surveys using methods of sociology and more recently, in collaboration with Gella Fishman, it has also incorporated the exhaustive elucidation and interpretation of archival material.

Fishman's concept of 'domain' has been particularly influential in promoting research into bilingualism and diglossia and related issues. For example, Parasher (1980) has applied it to the Indian bilingual cum diglossic situation. Similarly, Rubin (1968) applied it to the national bilingual situation in Paraguay.

Fishman has been criticized for having a European bias, and for threatening the fundamental assumptions of nation building. There is an internally consistent

response in Fishman's work to these criticisms: the preservation and promotion of the minority groups and their languages do not necessarily constitute a challenge to the larger societal groups. In fact, with planning awareness and appreciation of the potential contribution of all the elements, a rich and ultimately stronger social organization can be achieved.

Biography

Joshua A. Fishman was born in Philadelphia, PA, on July 18, 1926. He received B.S. and M.S. degrees from the University of Pennsylvania (1944–1948). He studied Yiddish, his mother tongue, at UCLA. He did his Ph.D. in social psychology from Colombia University (1953), and served as Educational Psychologist for the Jewish Education Committee of New York in 1951–1954. He directed research for the College Entrance Examination Board, while teaching sociology of language at City College (CUNY), in 1955–1958. He was Associate Professor of Human Relations and Psychology at the University of Pennsylvania (1958–1960); Professor of psychology and sociology at Yeshiva University, New York (1960); Dean of the Ferkauf Graduate School of Social Sciences (1966); served as academic vice president (1973–1975); and became professor emeritus in 1988. He held visiting appointments at ten major universities in the United States of America, Israel, and the Philippines, and fellowships at the Center for Advanced Study in the Behavioural Sciences (Stanford, CA), the East–West Center (Honlulu), the Institute for Advanced Study (Princeton, NJ), the Netherlands Institute for Advanced Study (Wassenar), and the Israel Institute for Advanced Study (Jerusalem). He has been editor of the *International Journal of the Sociology of Language* since its inception in 1973.

References

Fellman, J., and Joshua A. Fishman. 1977. Language planning in Israel: solving terminological problems. Language planning processes, ed. by J. Rubin et al. New York and Berlin: Mouton de Gruyter.

Fishman, G.S. 1991. Bibliographical inventory. The influence of language on culture and thought: essays in honor of Joshua A. Fishman's sixty-fifth birthday, ed. by R.L. Cooper and B. Spolsky. Berlin and New York: Mouton.

Fishman, Joshua A. 1964. Language maintenance and language shift as a field of inquiry. Linguistics 9. 32–70.

———. 1966. Language loyalty in the United States.

———. 1967. Bilingualism with or without diglossia, diglossia with or without bilingualism. Journal of Social Issues 23. 2. 29–38.

———. 1972. Language and nationalism: two integrative essays. Rowley, Mass: Newbury House Publishers (1973, ©1972).

———. 1989. Language and ethnicity in minority sociolinguistic perspective. Clevedon, Avon, England; Philadelphia: Multilingual Matters Ltd., ©1989.

———. 1968. Readings in the sociology of language. The Hague: Mouton.

———. 1985. The rise and fall of the ethnic revival: perspectives on language and ethnicity. Berlin, New York and Amsterdam: Mouton.

———. 1991. Yiddish: turning to life. Amsterdam: John Benjamins.

———. (ed.) 1999. Handbook of language and ethnic identity. New York: Oxford University Press.

Parasher, S.V. 1980. Mother-tongue-English diglossia: a case-study of educated Indian bilinguals' language use. Anthropological Linguistics 22(4).

Rubin, J. 1968. National bilingualism in Paraguay. The Hague: Mouton.

SHREE VALLABH PARASHER

See also **Diglossia; Language Planning; Sociolinguistics**

Focality

The terms focus, focal, and focality are used in two different domains: phonetics and syntax/semantics.

In phonetics, focus is understood as a specific stress or pitch; in a narrow sense, focal stress means the (typically falling, or rising–falling) sentence stress, placed in the unmarked (normal) case at the end of the sentence, and is distinguished from contrastive (typically rising) pitch that is often used at the beginning of a sentence: in the following examples, the latter is written in italics, the former in capital letters (with examples of typical preceding co-text segments closed in parentheses):

(1) (Mary visited us last week.) *Paul* was here YESTERDAY.

(2) (As a Christmas present, Jim got a picture from me.) *For Martin* his mother has brought a new BOOK.

(3) (Last Saturday I went to an exhibition with Jane.) *On Monday* we visited an old CASTLE.

(4) (Jerry and Marilyn were there too.) *Him* she SAW (, but she did not recognize Marilyn).

The contrastive pitch has to be distinguished from the presence of a focus in a marked position, which is the case in (1′):

(1′) (Who was visiting you yesterday?) PAUL was here yesterday.

In the analysis of the sentence and its information structure, focus is understood as that part of the sentence referring to 'new' information, rather than to the 'given' information (corresponding to the topic of the sentence). Phonetic focus is carried by a part of the informational focus; cf. the following examples, in which normal intonation is assumed (with the stress at the end of the sentence) and the sign '/' marks the boundary between topic and focus):

(5) (When did your friends visit you?) Some of them were present/at our party last Friday.

(6) Many men/read few books. —Few books are read/by many men.

(7) Mary was writing her dissertation/on the weekends. —On the weekends, Mary was writing her dissertation.

As the last two examples show, the articulation of the sentence in its topic and focus is semantically relevant. The passivization may be understood as triggering a word-order variation, for which otherwise English allows only in specific syntactic contexts, such as:

(8) They moved/from a village to a large industrial center. —They moved to a large industrial center/from a village.

(9) We went/by car to a lake. —We went to a lake/by car.

In many other languages (Latin, German, Czech, and most other Slavic languages), the word order is much freer than in English, so that e.g. an object belonging to the topic can precede the subject:

(10) Das Buch hat mir mein Vater gegeben (lit.: the book has me my father given—'My father gave me the book').

While a definite noun group included in the topic prototypically triggers a presupposition of the existence of an entity that it refers to (more precisely, to the communicative availability of the referent), noun groups contained in the focus are not connected with such a presupposition (a presupposition is an assertion, the truth of which follows not only from the sentence triggering it but also from its negative counterpart), cf.:

(11) Mike's sister has visited that exhibition (presupposition: *Mike has a sister*).

(12) That exhibition was visited by Mike's sister (no such presupposition, the existence of Mike's sister follows only from the positive sentence, but not from its negative counterpart).

It is more exact not to work only with the dichotomy of topic and focus, but to describe the information structure of the sentence on the basis of an opposition of contextually bound and nonbound items. This allows for a more subtle classification, cf. the following examples, in which the indices b and n are used for contextually bound and nonbound items, respectively.

(13) (A young pair was present there.) I_b recognized$_n$ only$_n$ him$_n$.

(14) (I prefer fiction written by a not much known lady from Edinburgh.) This$_b$ author's$_b$ books$_b$ have$_n$ a specific$_n$ flavour$_n$ for me$_b$.

In the unmarked case, nonbound items belong to the focus of the sentence and bound ones are included in its topic; exceptions concern bound items embedded in the focus (such as *for me*, if understood as a modifier of *flavour* in (14)) and nonbound items in the topic.

Note that the opposition of contextual boundness is understood as a pair of grammatical values, which corresponds to the cognitive notions of 'given' and 'new' information in the prototypical case, but not always; e.g. *Him* in (13) as a stressed pronoun belongs to the focus of the sentence and is nonbound, although its pronominal (anaphoric) character is connected with the fact that it refers to an entity known to the speaker (at least in some sense) and that is assumed to be known also to the hearer.

References

Bosch, P., and R. van der Sandt (eds.) 1998. Focus. Linguistic, cognitive and computational perspectives. Cambridge: Cambridge University Press.

Hajičová, E., B.H. Partee, and P. Sgall. 1998. Topic-focus articulation, tripartite structures and semantic content. Dordrecht: Kluwer Academic Publishers.

Steedman, Mark. 2000. Information structure and the syntax–phonology interface. Linguistic Inquiry 31. 649–89.

EVA HAJICOVA

See also **Context; Definiteness; Prosody**

Forensic Linguistics

Forensic linguistics is the scientific study of language as applied to forensic purposes and contexts—the application of linguistic knowledge to legal problems. It is a new area of linguistics and a rapidly growing area of modern applied linguistics.

Recent research demonstrates increased interest and research in forensic linguistics. Overviews of language, law, and the legal process have been elaborated by many scholars, e.g. Levi (1994b). Work has continued on the language of the courtroom and on analysis of discourse in the language used in legal settings (Shuy 1993, 1998). Many specific studies and collections relating to linguistic applications to the law have been written or edited, one of the first and most significant being that of Levi and Graffam Walker (1990). Documentation of advances in forensic linguistics appears in the research of individual forensic linguists and in the bibliographic work of Levi (1994a) and of the University of Birmingham.

Recent milestones in forensic linguistics include the following: Judith Levi and Anne Graffam Walker organized and coordinated the 1985 Georgetown University conference on Language in the Judicial Process. In 1995, Bethany Dumas started *Language in the judicial process*, an electronic newsletter aimed at disseminating information on bibliography, organizations, courses and programs, and legal cases. During the 1990s, university courses on language and law were developed and presented around the world. The most important advances in the study of forensic linguistics were the University of Birmingham's initiation of *Forensic Linguistics: The International Journal of Speech, Language and the Law* in 1994 and their founding of the International Association of Forensic Linguistics, which have since provided serious venues for the presentation of research that are more regular, unified, and formal than ever before.

The classification of areas in forensic linguistics is evolving as the field develops, usually following those of the structure and function of language. The research and casework of forensic linguists presently define the field in a taxonomy that includes the following areas: auditory and acoustic phonetics, semantics, discourse and pragmatics, stylistics, language of the law and of the courtroom, and interpretation and translation.

Auditory phonetics is the study of speech sounds based on what is heard and interpreted by the human listener (the aural–perceptual characteristics of language),

although studies in forensic phonetics often use both auditory and acoustic methods of analysis. The primary areas of auditory research in forensic phonetics are speaker discrimination and identification by victims and witnesses; voice perception, discrimination, imitation, and disguise; and identification of group characteristics of speakers, including first-language interference, regional or social accent and dialect, and speaker age. An introduction to the use of auditory (and acoustic) phonetics in forensic examinations is French (1994).

Acoustic phonetics is the study of the physical characteristics of speech sounds as they leave the speaker, move into the air, and gradually dissipate. The acoustic analysis of speech sounds requires laboratory observation with instruments and specialized computer analysis.

Vowels and consonants are presented as waveforms having three features: amplitude, corresponding to loudness; frequency of complete repetitions, corresponding to high/low pitch; and complexity, corresponding to the periodic waves of a simple sound, or to a mixture of waves from a complex sound, like a vowel sound. The waveforms are presented visually as a spectrogram. See Hollien (1990) for the forensic application of acoustic phonetics.

The primary area of acoustic analysis in forensic phonetics is speaker identification, but many studies have also been carried out to identify group characteristics of speakers, including their physical height and weight; regional, social, or language group; voice and accent disguise; effect of intoxication on speech; and technical aspects of speech samples and recordings.

Semantics is the study of meaning as expressed in words, phrases, sentences, or texts. The focus of semantic analysis in forensic contexts is on the comprehensibility and interpretation of language that is difficult to understand. Some studies combine semantic and pragmatic approaches to meaning interpretation. A thorough introduction to this area can be found in Solan (1999). The point of view that expert linguists have no role to play in helping judges interpret statutes is presented by Murphy (1998).

Primary areas of research in forensic semantics are: interpretation of words, phrases, sentences, and texts; ambiguity in texts and laws; and interpretation of meaning in spoken discourse, such as reading of rights and police warnings, police interviews, and jury instructions.

Analysis of discourse is the study of units of language that are larger than the sentence, such as narratives and conversations. Discourse in spoken and written language can take many forms, especially in conversations tied to specific social contexts. Analysis of a speaker's intended meaning in actual language use is the study of pragmatics. Pragmatics is important for forensic purposes because speakers and writers do not always directly match the words they use with the meaning they intend to convey. This leaves the speaker's intended meaning more open to interpretation by the listener, sometimes resulting in miscommunication.

The linguist most closely associated with developing forensic discourse analysis for a broad range of cases is Roger Shuy. Some of his cases are documented in *Language crimes: the use and abuse of language evidence in the courtroom* (Shuy 1993) and in *The language of confessions, interrogation and deception* (Shuy 1998).

Primary areas of research in forensic discourse and pragmatics include: analysis of spoken and written language; study of the discourse of specific contexts, such as dictation, conversations, and hearings; the language of the courtroom, i.e. of lawyers, clients, questioning, and jury instructions; and the language of specific speech acts, such as threats, promises, and warnings.

The focus of forensic stylistics is author identification of questioned writings. While the methods were developed mainly for literary purposes, applications of stylistic analysis to forensic authorship problems have become very common. In criminal and civil cases with questioned writings, there is often a need to determine whether one author wrote all the writings in a questioned set, whether one of a number of possible authors authored the questioned writing, or whether a single suspect-author can be eliminated or identified as the writer.

Linguistic stylistics uses two approaches to authorship identification: qualitative and quantitative. The work is qualitative when features of writing are identified and then described as being characteristic of an author. The work is quantitative when certain indicators are identified and then measured in some way, e.g. their relative frequency of occurrence in a given set of writings. Certain quantitative methods are referred to as stylometry. Qualitative and quantitative methods complement one another and are often used together to identify, describe, and measure the presence or absence of style markers in questioned and known writings.

Examples of descriptive (qualitative) style markers from two separate cases appear in Figure 1.

Another recent development in forensic authorship identification is the application of stylistic analysis to computer programming. Researchers (Krsul 1994) and security analysts, working on proprietary issues and virus sources, have identified various style indicators as indicative of authorship in programming code, e.g. variable names, layout, upper/lowercase letters, placement of comments, debugging symbols, line length, and ratio of white lines to code lines.

Linguistic Unit	In QUESTIONED Writing	In KNOWN writing
Same-Author Case A: *confidentiality*	confidentuality	-Confidentuality
Same-Author Case A: *I'll see you.*	A SEEYOU	a See you later
Different-Author Case B: Spanish *hice*	_ise	hice
Different-Author Case B: Spanish *estaba*	_estaba.	ESTAVA

Figure 1. Style-marker examples for spelling and word formation.

Primary areas of research in forensic stylistics relate to the descriptive and quantitative methods of authorship identification, dialect evidenced in written language, the questioned time and occasion of writing, stylometry, statistical methods of analysis, use of corpus linguistics in forensic analysis, and computer programs for the analysis of style. A relatively complete survey of forensic stylistics can be found in McMenamin (2002).

Perhaps the most important development in the language of the law was that begun by David Melinkoff in his 1963 book, *Language of the law*. Melinkoff pressed for clarity and brevity in the law, which extended to a later movement to simplify the language of laws, insurance policies, and consumer literature. Systematic analysis of legal language is carried on today with, for example, Tiersma's (1999) emphasis on plain legal language, language rights, and the pragmatics of written legal language as well as of courtroom language.

The courtroom personae who speak are witnesses, lawyers, and judges. Research related especially to the discourse and pragmatic use of courtroom language has made this one area of forensic linguistics that can significantly affect case outcomes. Studies of courtroom language are analyses of the language of all the players: the language of witnesses (witness examination, victims, children, men vs. women); the language of lawyers (trial language, legal debate, closing arguments); and the language of judges (trial participation, jury instructions).

Interpreting is a complex skill and especially difficult in forensic contexts. Interpretation studies such as those in Cooke et al. (1999) focus on interpretation tasks specific to questions and answers in testimony, the perceived role of the interpreter, interpreter education, the right to interpretation, etc. Forensic and academic scholars are giving more and more attention to the theory and practice of interpretation and translation, especially in these areas, and in others such as pretrial interpreting, courtroom interpretation, interpretation with cultural and dialect differences, questioning in interpreted testimony, and the absence of interpretation.

Translating in the legal context requires much more than a literal, word-for-word matchup of two languages. Good translations are constrained by the intended meaning of the writer, the new text created by the translator, and the meaning given the translated text by the reader.

Areas associated with forensic linguistics include document examination and aspects of psycholinguistics. The examination of questioned documents relies on the scientific study of the physical evidence of a document. Forensic document examiners look to the writing instrument, the writing surface, and the writer for physical traces that assist in uncovering the history of a document. The document examiner observes features of handwriting (letter size, formation, and relative proportions; letter slant, spacing, pressure, line quality, connecting strokes, etc.) and of typing (typeface or font style, spacing between letters and lines, association of a document with a particular machine or type of typewriter, printer, or copier).

Previous studies indicate that examination for authorship determination includes numerous characteristics of writing style (punctuation, spelling, abbreviations, forms of dates, etc.), but newer studies are drawing a sharper line between forensic document examination and the elements of writing style studied in linguistic stylistics.

Psycholinguistics is the academic discipline that integrates the study of psychology, linguistics, and cognitive science. The acquisition, comprehension, and production of language are studied in various related ways: cognitively, neurologically, and conceptually. However, this is not what has been referred to as 'psycholinguistics' in forensic contexts. The late Murray S. Miron developed the so-called 'psycholinguistic approach', which examines written or spoken language to profile the origins, background, and psychology of the originator, especially in the context of threats. It is difficult for a nonpsychologist (e.g. a judge or jury) to understand and evaluate the nexus made between a written-language threat and the diagnostic profile of the writer, and the practice of 'psycholinguistics' can be risky and may lead to mistakes. This kind of 'psycholinguistics' is not commonly regarded as a subdiscipline of forensic linguistics.

References

Bryant, M. 1930, 1962. English in the law courts: the part that articles, prepositions and conjunctions play in legal decisions. New York: Frederick Ungar.

Cooke, M., D. Eades, and S. Hale. 1999. 'Introduction,' to special issue on legal interpreting, with eight other articles. Forensic Linguistics 6(1). 1–5ff.

Danet, B. 1980. Language in the legal process. Law & Society Review 14(3). 445–564.

French, P. 1994. An overview of forensic phonetics with particular reference to speaker identification. Forensic Linguistics 1(2). 169–81.

Hollien, H. 1990. The acoustics of crime: the new science of forensic phonetics. New York: Plenum.

Krsul, I. 1994. Authorship analysis: identifying the author of a program. Technical Report CSD-TR-94-030, Department of Computer Science, Purdue University.

Levi, J.N. 1982. Linguistics, language, and law: a topical bibliography. Bloomington: Indiana University Linguistics Club.

———. 1994a. Language and law: a bibliographic guide to social science research in the USA. Chicago: University of Chicago Press and Washington, DC: American Bar Association, Teaching Resource Bulletin No. 4.

———. 1994b. Languages as evidence: the linguist as expert witness in North American courts. Forensic Linguistics 1. 1–26.

Levi, J.N., and A. Graffam Walker (eds.) 1990. Language in the judicial process. New York: Plenum Press.

McMenamin, G.R. 2002. Forensic linguistics. Boca Raton: CRC Press.

Melinkoff, D. 1963. The language of the law. Boston: Little Brown.

Murphy, H.F. 1998. Linguistics and law: an overview of forensic linguistics. Journal of Law, Intellectual Property, and Technology 1. 1–20.

O'Barr, W. 1982. Linguistic evidence, language, power and strategy in the courtroom, New York: Academic Press.

Shuy, R.W. 1984. Linguistics in other professions. Annual Review of Anthropology 13. 419–45.

———. 1986. Language and the law. Annual Review of Applied Linguistics 7. 50–63.

———. 1993. Language crimes: the use and abuse of language evidence in the courtroom. Oxford: Blackwell.

———. 1998. The language of confessions, interrogation and deception. Los Angeles: Sage Publications.

Solan, L.M. 1999. Can the legal system use experts on meaning? Tennessee Law Review 66 (summer) 1167.

Tiersma, P.M. 1999. Legal language. Chicago: University of Chicago Press.

Wetter, J.G. 1960. The styles of appellate judicial opinions. Leyden: A.W. Sythoff.

Electronic Resources

Forensic Linguistics: The International Journal of Speech, Language and the Law: http://www.bham.ac.uk/forensic linguistics/

International Association for Forensic Phonetics: http://www. iafp.net/

Language in the Judicial Process: http://www.outreach.utk. edu/ljp/

Plain Language—USA http://www.plainlanguage.gov/

Plain Language—UK http://www.plainenglish.co.uk/

GERALD MCMENAMIN

See also **Discourse Analysis; Psycholinguistics; Semantics**

France

This article deals with the linguistic situation of France, and with the French language in its international context. Since 1992 the French constitution has stated that the language of the republic is French. As a major western European state (population 59 million) and former imperial power, France has a long history of standardizing and promoting French, although there are other languages within its borders.

The French Language

French is classified as a Gallo-Romance language. The Romans who colonized Gaul in the first century BC had a profound and lasting influence on the institutions, culture and language of the country that we know as France. Of the period prior to Roman domination, only a few traces remain of the Gaullish (Celtic) substratum—everyday words e.g. for trees *if* 'yew tree', or place names, e.g. the *dun* ('fort') ending found in *Verdun*. The Romans spread their language (or languages, from classical written Latin, to a later more popular and largely spoken form) so successfully that when Germanic tribes invaded in the fifth century AD, their languages influenced the Latinate language in use (e.g. in vocabulary items such as *guerre* from *werra*) but did not supplant it. Evidence for the emergence of a recognizably distinct French language (or set of dialects) is found in the bilingual Strasbourg Oaths (842 AD). The dialects of the north, the *langues d'oïl*, showed more Germanic influence, as against the Romance *langues d'oc* in the south. (This division is still reflected today in the existence of three main linguistic zones, the north being divided from the south by a line running east from Poitiers, with an intermediate *franco-provençal* triangle fanning out from the region of the Alps.) Among the major dialects of the north with a written form, it was the language of the Ile de France (around Paris) which was set to become the high prestige variety. So-called 'Old French' of the twelfth and thirteenth centuries had a wide range of vocalic sounds and had reduced to two the Latin system of cases. During the fourteenth century case endings disappear, with a corresponding reliance on determiners, the vocalic system is simplified and we find that the final consonants have been dropped. Through the Edict of Villers-Cotterêts of 1539, French replaced Latin as the official language for administrative and legal documents. Ten years later the writer Joachim du Bellay in his *Défense et illustration de la langue française* advocated the use of French in literature; du Bellay and other writers contributed greatly to the enrichment of the French language, although the Renaissance also brought in its wake an influx of Italian loan words. The seventeenth century saw the codification of the 'standard' language, with the *Académie Française* being set up in

1635. Nevertheless, there were still significant numbers of speakers of other languages by the mid-eighteenth century when the revolutionaires tried to ensure the development of a unified state. A report by the Abbé Grégoire estimated that no more than three million (about a fifth of the population) mastered standard French. From the late nineteenth century onwards, standard (Parisian) French was further promoted by free, universal schooling (from 1880 on by Jules Ferry), conscription in World War 1 and the mobility of civil servants. Today, facilitated also by the media, standard French dominates across the whole of France, sometimes alongside a regional form of French and sometimes alongside a regional language or dialect. Even apart from dialectal variation, there is a good deal of sociostylistic (register) variation. While the efforts of bodies such as the *Académie* to conserve the language have applied mainly to a formal, usually written, register, the spoken language demonstrates certain syntactic differences (e.g. deviation from canonic Subject–Word–Object word-order, dropping of pre-verbal negative particle *ne*) and characteristic lexical features (a rich range of colloquial vocabulary for everyday items (*bouquin-livre* 'book') with many abbrieviated forms (*apéro-apéritif* 'before-dinner drink'). So-called *français jeune* refers to the French spoken by some young people, characterized by a mix of colloquial French, 'slang' vocabulary such as *le verlan* based on the inversion of syllables with subsequent modifications (thus *laisse béton—laisse tomber* 'leave it', *meuf —femme* 'woman') and borrowings from various languages. Borrowings from English are plentiful in all registers of contemporary French, hence recent attempts to legislate in favor of the protection of French, from the Bas-Lauriol law of 1975 to the controversial *loi Toubon* passed in 1994. A debate that had rumbled on for years about orthographic reform was finally settled in 1991 with a compromise solution, with a small number of spelling rationalizations being recommended but not imposed. Apart from the *Académie* which still pronounces (usually prescriptively) on language matters, there are currently various commissions devoted to protecting French, such as the *Conseil supérieur de la langue française* which promotes 'good usage', and the *Délégation générale à la langue française et aux langues de France* part of whose job is to liaise between the terminological commissions set up by each government ministry to stem the influx of English words, as well as to promote French world-wide.

The Other Languages of France

The regional languages that the revolutionaries and educationists had tried to suppress began to fight back in the mid twentieth century. Their modest revival was eventually made possible by the place won for them within the French education system, and by a shift in government thinking initiated by the socialist president Mitterand's recognition of the 'right to difference' in 1981. The *loi Falloux* of 1851 had stipulated that French was the only language to be used in schools, but 100 years later the *loi Deixonne* allowed schools to give a few hours optional teaching a week in those regional languages that were not the language of a foreign power (Corsican, considered a dialect of Italian, was only added in 1974). Since 1981 government policy has become gradually less hostile to the regional languages, which have been boosted by the *écoles associatives* in which they can be the medium of instruction. Today, the following are recognized as 'regional languages': three Romance languages, Occitan, Corsican (close to Italian) and Catalan (an official language of Catalonia in Spain); two Germanic languages (Alsatian in Alsace and part of Lorraine) and Flemish for a small number of speakers near the Belgian border; and one Celtic language (Breton) and the non-Indo-European language Basque (also spoken on the Spanish side of the border). It is difficult to get an objective estimate of the numbers of speakers, or even of the health of these languages (but see Judge 2000, Sanders 1993 and the publications of the European Union on the Lesser Used Languages of Europe EBLUL).

Basque (approximately 80,000 speakers) has been boosted by the *ikastolak* (*Basque language schools*), which have received state funding for primary and secondary levels since 1994. The first school opened its doors in 1969 to five pupils and by 1997 there were 1630 pupils receiving nursery, primary or secondary education through the medium of Basque. There are four dialects of **Breton** and probably about 500,000 speakers. In 1997, 1,751 pupils sat for the French school-leaving exam (*baccaulauréat*) having followed a complete cursus through the *diwan* (Breton schools) created 20 years earlier, and had the possibility of taking some courses in Breton e.g. at the University of Rennes. There are an estimated 100,000 speakers of **Catalan** in the Languedoc-Roussillon area of France, who are at once encouraged and overshadowed by the support given to the language by the provincial government of Catalonia on the Spanish side of the Pyrenees. Catalan schools started in 1976 and it is now possible to go up to doctoral level at the University of Perpignan. The future of **Corsican** (approximately 200,000 speakers) is linked to on-going discussions about the status of the island, including the possibility of a regional assembly and of the increased use of Corsican in schools, many inhabitants being bilingual in French and

Corsican. In a 1982 survey by the French agency INSEE, 96% of Corsicans on the island understood the language and about 86% claimed to speak. It is possible to study Corsican up to postgraduate level (University of Corti), and there has been a strong literary revival. In some respects the Germanic languages of north-east France fare less well in the education system (and in the media) than other regional languages. As **Flemish** and **Alsatian** were classified as official languages of a neighbouring state, no support was given under the terms of the *loi Deixonne*. A current estimate of 80,000 speakers of Flemish may be optimistic. Alsatian is treated as a spoken dialect for which the written form is German, and the high number of those represented as studying this 'regional language' is misleading, as it includes pupils taking German, and for Alsatian itself there are no schools, only *classes associatives*. **Occitan** should be the giant of the regional languages, covering as it does 31 *départments* in the southern half of France. However, various factors such as migration to and from the area, and dialectal differences between Gascon and Northern and Southern Occitan (the latter dividing further into Provençal and Langdocien), mean that the survival of this language group, with its rich medieval literary heritage, cannot be taken for granted. The orthography proposed by the nineteenth century poet Mistral and his association, the *Félibrige*, was not necessarily suitable for Occitan in the twentieth century and a compromise form has now been reached. A flowering of theatre and song in the second half of the twentieth century was accompanied by a degree of political activism. There are now over 30 *calendretas* (schools), and 12,532 children were taught Occitan in the school year 1997–8. Overall, there may be up to 10 million people who understand the *langues d'oc*, and two million who practise them. The **langues d'oïl**, the dialects of northern French, were treated as 'patois' and given little consideration for many years, but a number of studies have recently demonstrated the tenacity of *picard* or the *chtimi* dialect of the Lille area.

Beside the indigenous languages of metropolitan France, there are other languages that are worthy of mention. A report by Bernard Cerquiglini in 1999 identified 75 languages in 'greater' France. Firstly, the other languages of the overseas departments of France (e.g. the Polynesian and Melanesian languages of the French Pacific territories) are technically 'languages of France'. Secondly, so are the non-territorial languages (Romany languages, Yiddish) and the immigrant languages of metropolitan France. The latter are the result of successive waves of migration—Polish, Italian, Spanish, Portuguese, Chinese, and speakers of sub-Saharan African languages, to mention only a few.

The largest groups currently originate from North Africa: thus, there are estimated to be approximately 1.5 million speakers of Arabic and 0.5 million Berber speakers in France. There have been some attempts to provide instruction for Arabic heritage pupils, but so far these have partly foundered on such issues as which dialect to teach (classical, Algerian, etc.). To sum up, levels of educational provision vary from language to language, as do levels of competence. As for most languages with no official backing, the regional languages have suffered from dialect fragmentation, and the norm adopted has not pleased all activists. Those now being educated in the *écoles associatives* may be in a position of having parents who do not speak the language and grand-parents who speak a different dialect of it; few will have the chance to speak the language in the work-place. Within the European Union, French is one of the two major working languages, although constantly under pressure from English. After some delay France became a signatory to the European Union's Charter for Regional or Minority Languages but did not ratify it, on the grounds that it is incompatible with the constitution which, holding all citizens to be equal, does not recognize minorities. However, the title of the *Délégation générale à la langue française* has been amended to include the 'languages of France' which may be a hopeful sign for the regional languages.

The French-Speaking World (la Francophonie)

France's first wave of colonial expansion in the sixteenth century took the French language to the Americas, and its second during the nineteenth and the twentieth centuries to North and sub-Saharan Africa, the Pacific and parts of Asia. Today a reminder of this imperial past remains in the use of French in France's former territories (Québec) as well as in her overseas departments (such as the French Antilles) or territories (e.g. French Polynesia). Depending on the country or area, French may be the only official language (as in Québec), one of two or more official languages (cf. French and English in New Brunswick or the same two languages in Cameroun), one of several official languages of a multi-lingual state (Belgium or Switzerland), or it may be a major language of education and business without being an official language (e.g. Morocco or Madagascar). It is sometimes claimed that the future of French lies with its use outside France and that will be, for example, more speakers of French in Africa than in France. However, this depends on the definition of a 'French-speaker', not all of those who use French as a vehicular language being fluent speakers, and such varieties as *le français*

populaire d'Abijan differing substantially from standard metropolitan French. What follows is an attempt to estimate the numbers of French-speakers in illustrative situations from around the world.

Europe To the 59 million citizens of metropolitan France must be added about 6.5 million speakers of French in Belgium, Switzerland, Monaco, Luxembourg, Andorra and small pockets such as the *Val d'Aosta* in Italy. 4.5 million of these are in Belgium where French is one of three official languages, the other two being Flemish (i.e. Dutch) and German. The original superior status accorded to French when Belgium became a nation in 1830 gave rise to a succession of protests, which eventually led to the establishment of a trilingual state, with parallel organizations for the French and (now majority) Flemish speaking populations. Thus, apart from the small German-speaking community near the German border, the country is divided into three linguistic entities: the French-speaking community (which includes Belgian French and the Romance *wallon* dialect), the Flemish-speaking community, and Brussels, accorded special bilingual status. Educated Belgian and Parisian French are very close, with some particular features characterizing Belgian French such as the use of the numerals *septante, octante* etc. for standard French *soixante-dix* ('sixty') and *quatre-vingts* (eighty).

The Americas French and English are the official languages of federal Canada, though each province determines its linguistic status. The majority of French-speakers are in Québec (population 6.5 million) where French is the only official language, but there are also French speakers elsewhere, notably in the Maritime provinces, Ontario, Manitoba and Saskatchewa. *Québécois* French retains traces of the sixteenth-century regional speech of the original migrants, as in the characteristic pronunciation of /mwe/ for *moi* ('me/I'). The linguistic range extends from educated *québécois* French, which may show certain differences, largely in pronunciation and lexis, from metropolitan French, to a variety of rural and urban dialects—in the latter case most notably the urban working-class *joual* of Montreal. In the last half century, French in Quebec has benefited from strenuous provincial government efforts to promote its use through linguistic legislation and the setting up of bodies such as the *Office de la langue française*.

French-speaking emigrants deported by the English from the Arcadian area of the Maritimes took their language to New England and to Louisiana where French was further bolstered by French Canadian economic migrants (in New England) and by French and Creole-speaking migrants from the Caribbean (in the case of Louisiana). In the Caribbean, both French and French-based Creole are spoken in independent Haiti and in the French departments of Guadeloupe, Martinique and Guyane, but only French is recognized as the official language in the latter.

Africa French is widely used as a vehicular language in the former French colonies in sub-Saharan Africa. The situation is fluid, but currently the following are among the countries in which French is an official language but co-exists in daily use with African languages: Burkina Faso, Congo, Guinea, Ivory Coast, Mali, Niger, Senegal. In North Africa (Tunisia, Algeria, Morocco), there has been a policy of 'arabization', but French is still widely used. In the Indian Ocean, French is used in education in Mauritius (official language: English) and Madagascar (official language: Malagasy) and Réunion (a department of France, where Creole co-exists with the official language French).

Since some speakers have French as a second or third language, estimating the number of speakers worldwide is not easy, but the estimate of 80–100 million with French as a mother-tongue is sometimes given. A great strength of French is that it is present on every continent, and another is that educated speakers from around the world speak a mutually comprehensible form of French which approximates to the Parisian norm. An international 'francophone summit' meets every two years to promote cooperation among French-speaking countries. In addition to France, certain other countries now play an important role in promoting French worldwide. In certain areas e.g. terminology banks, or the 'feminizing' of the language (particularly for job titles), Francophone countries such as Canada have acted more effectively than has France herself. French is indubitably the vehicle for a vibrant *francophone* culture, in literature, cinema and media in many countries.

References

Ayres-Bennett, W., and J. Carruthers, with R. Temple. 2001. Studies in the Modern French language: Problems and perspectives. London: Longman.

Ball, R. 1997. The French-speaking world. London and New York: Routledge.

Judge, A. 2000. France: 'one state, one nation, one language?' Language and nationalism in Europe, ed. by S. Barbour and C. Carmichael. Oxford: Oxford University Press.

Lodge, A., N. Armstrong, Y. Ellis and J. Shelton. 1997. Exploring the French language. London: Arnold.

Sanders, C. (ed.) 1993. French today. Cambridge: Cambridge University Press.

Walter, H. 1994. French inside out. London and New York: Routledge.

CAROL SANDERS

See also **Indo-European 4: Romance**

French Language

French is one of the major languages of Europe, spoken as a first language not only in France (where it originated) but also in parts of Belgium, Switzerland, Luxembourg, and a small region of Italy (the Aosta Valley). In addition, as a result of colonization, French is spoken in many other parts of the world. In Africa, it is the official language of 18 countries, which occupy a good half of the land area of Africa, from Mauritania in the west to the Central African Republic to Chad, Mali, and Senegal in the east to Congo, Zaire, Rwanda, and Burundi in the south. It is also spoken and understood by a considerable section of the population in the three Arab countries of North Africa that constitute what is called the Maghreb (Morocco, Algeria, and Tunisia). Across the Atlantic, in North America, it is one of the two official languages of Canada and is the first language for many Canadians in Québec and other provinces of Eastern Canada. There are vestiges of French in the United States as well, particularly in Louisiana, Missouri, and parts of New England, and in the Caribbean islands of Martinique and Guadeloupe (which are still dependencies of France), as well as the Republic of Haiti. On the continent of Asia, it was spoken by many in former French Indo-China and still retains its prestige in Laos and Kampuchea (former Cambodia), but is much less used in Vietnam than formerly. There are many islands in the Indian Ocean and Oceania where French is still spoken: Madagascar and smaller islands nearby, and New Caledonia, French Polynesia (including Tahiti), New Hebrides, and various smaller islands. In all, there are about 88 million francophones (speakers of French) in the world. After English, it is the second international language, used by institutions such as the United Nations and UNESCO (United Nations Educational, Scientific, and Cultural Organization).

Although French is the official language of France and, since the beginning of the twentieth century the media have been accessible everywhere, not everyone in France speaks French. First of all, there are still other languages spoken in France: e.g. Basque in the southeast (on the border with Spain), Breton in the west, Flemish in the north (on the border with Belgium), and Alsatian and Lorrain in the east (on the border with Germany). In addition, there is Occitan (including Provençal) in the south, Franco-Provençal (spoken in France, Switzerland, and Italy), as well as many small local languages spoken all through France. All these have their effect on French. In Southern French (Occitan area), for example, mute (written) *e*'s tend to be sounded, while they are silent in Standard French. Alsatian French shares many features of pronunciation with the Germanic dialect spoken in the region.

The same is true of French spoken outside France. The French spoken in the rest of Europe tends to differ from the French of France most notably in pronunciation and somewhat in vocabulary. The numbers 70, 80, and 90, for example, have different names in Belgium and Switzerland than in France. But it is the French outside of Europe that shows the most diversity—e.g. québécois in Canada, Cajun in Louisiana, the various types of French in former Indo-China, the Indian Ocean, French Polynesia, and Africa all have developed their own specific traits of the language. As one example, in Canadian French, stress placement (word accent), intonation (inflection of the voice), and the vowel system are not the same as in Standard French, nor is the vocabulary. As for French creoles (found in Louisiana, French Guyana, the Caribbean, and the Indian Ocean), they are not intelligible to speakers of French. Their vocabulary is based on that of French, but they are entirely new languages with their own particular morphology and syntax.

French has a variety of roles in the linguistic repertoire of its users. It can be a native language (for example, in France, Belgium, Canada, Luxembourg, and Switzerland). In many countries, it is an (or even the) official language used for governmental and other purposes (in the nations just mentioned; also in Haiti, the Comoros, and in many former French and Belgian colonies in sub-Saharan Africa). In still others, it functions as a *lingua franca* (a language of communication by those whose own languages are very different from each other), for example, in sub-Saharan Africa and in New Caledonia. For many people, it is their second or third language in contexts where there may be many other languages around; in such cases, it is highly prestigious to speak French and thus only the well-educated elite do (in French Polynesia and in Creole-speaking areas such as Haiti, French Guyana, Mauritius, some of the Caribbean Islands, and the Seychelles). It is also, for some, a language of culture (in Lebanon and the Maghreb). In many countries outside *Francophonie* (those countries where French is recognized as a means of communication), French remains—after English—a preferred foreign language learned at school.

357

History of French: Old and Middle French

French is a descendant of the Vulgar Latin (common spoken language) brought by Roman settlers to the area occupied by modern-day France in the last century BC, after it was conquered by Julius Caesar in the Gallic Wars. It became the spoken language of Gaul, as the region was then known, and remained so even after the fifth-century invasions by Germanic tribes. But this Gallo-Romance way of speaking diverged increasingly from Latin, as recognized by the Council of Tours (813), which recommended not using Latin in preaching because the common people could not understand it. Moreover, by the end of the ninth century, the northern dialects were most influenced by the Germanic speech of the dominant Franks, and the southern dialects remained closer to Latin. This resulted in the emergence of two separate languages: *langue d'oïl* in the north and *langue d'oc* in the south (*oïl* and *oc* were the words used for 'yes' in the two languages). Modern French is a descendant of *langue d'oïl*. The oldest document in *langue d'oïl* is the *Strasbourg oaths* (843), which marks the beginning of the Old French period (ninth to fourteenth centuries). Latin being virtually the only written medium of the early middle ages, documents in Old French are rare. Literary texts appear first in *langue d'oc*, the language of the *troubadours*, at the end of the eleventh century, a hundred years before their counterparts in Old French.

In the tenth century, French kings made their court in Paris and, as the royal domain increased, so too did the prestige of the dialect they spoke (called *francien*). In the twelfth century, Parisian speech was already a model to be emulated; by the end of the Old French period, the King's French was in the process of becoming the written norm in France.

In the Middle French period (fourteenth to sixteenth centuries), the use of French continued to expand. In 1539, a royal edict banned Latin from court proceedings and legal deeds and replaced it with the French spoken at the time. French was also promoted as a vehicle of learning and culture at the expense of other dialects, which were relegated to the status of *patois* (local vernaculars).

In 1549, Joachim Du Bellay's *Defense and illustration of the French language* maintained that French was not inferior to the classical languages (Latin and Greek) and should be used for literary works. During this period, the King's dialect began to be codified and, with the advent of printing, attempts were made to simplify and regularize the orthography.

Classical French

The Classical French period (seventeenth to eighteenth centuries) witnessed further efforts to standardize the language. In 1635, the French Academy was founded to give 'firm rules' to the language, to make it 'pure and eloquent', and to compose a grammar and a dictionary. Claude Vaugelas, in his *Remarks about the French language* (1647), established that the linguistic norm should be based on 'good usage', i.e. the French spoken at the Court and used by the best writers. During the Classical period, French acquired great prestige abroad. It was the international language used in European courts and became the medium of diplomacy. Such was the fame of French that competitions were held to extol its virtues. In 1784, Antoine Rivarol won first prize from the Academy of Berlin for his 'Discourse on the universality of the French language'. Great writers also contributed to the preeminence of French and the development of the literary language.

Modern French

By the time of the French Revolution in 1789, French had spread to the bourgeoisie (middle class) in urban areas; but outside the Paris region, the *patois* were still very much the vernacular of the people. In 1790, Abbé Grégoire conducted the first survey about the use of French and found that 45% of the population of France were not able to speak French. The Revolutionaries held that linguistic uniformity was essential to the cohesion of the nation and equated the *patois* with backwardness and antirepublican sentiments. They wanted, therefore, to exterminate the *patois* and to have French taught at school to every citizen. These ideas were not implemented until almost a century later, during the Third Republic. Through laws devised by Jules Ferry and passed in 1881–1886, primary school education was made free, secular, and compulsory. The teaching of French was central to the curriculum, and the use of any other language at school, inside or outside of the classroom, was banned and punished. In addition, railways, roads and cities, military service, mass media (newspapers, radio, television)—all contributed to impose French as the only language of the nation in the twentieth century. In 1992, an amendment to the Constitution officially proclaimed French to be 'the language of the Republic'. At the same time, there has been a more tolerant attitude toward regional languages and regional differences. In 1951, the educational policy was relaxed to allow some regional languages, e.g. Basque, Occitan, and Breton, to be taught in French schools (the Deixonne law). By then, however, there were hardly any school-age native speakers of *patois* left and the demise of these languages is all but irreversible.

Protecting the French Language

With the rise of Anglo-American influence in the years following World War II in international commerce,

science, technology, and especially popular culture, a strong view has emerged in some French government and intellectual circles that French is under serious threat from English. While the proportion of English borrowings in everyday speech is in fact very low, concern for maintaining the purity of French has in recent times led to the creation of the General Delegation for the French Language (1989) and various ministerial committees charged with developing French vocabulary in economic and technical domains. The Bas-Lauriol (1975) and Toubon (1994) laws were passed to limit the use of foreign languages in the spheres of trade, employment, and services. Similar actions were taken in Québec where Anglo-American influence is overwhelming. The Charter of the French Language (1977) makes French the only official language in Québec and gives precedence to French in all areas of public life (education, work, economy).

Francophonie

In the 1960s, the notion of a francophone identity for users of French around the world became prevalent. African leaders suggested a new type of association, culturally and linguistically based, called *Francophonie*, which would replace the old colonial ties. The francophone movement was enthusiastically supported by France, Québec, and French-speaking Belgium. In France, the Supreme Council for Francophonie (founded in 1984) has contributed actively to the promotion of French outside of France. And since 1986, *Francophonie* summits have been organized every other year with about 40 participating countries.

Romance Languages

French is a Romance language because it has descended from the language spoken by Roman settlers as they went out to conquer neighboring lands and founded the Roman empire. The other Romance languages include Spanish, Catalan, Portuguese, Occitan, Italian, Rhaeto-Romance (spoken mostly in Switzerland), Dalmatian (spoken on the Dalmatian coast, in former Yugoslavia), Rumanian, and Sardinian. As descendants from Latin, the Romance languages have much in common, in particular their vocabulary. They also share other traits that they inherited from Vulgar Latin. Noun morphology was simplified in most of the Romance languages, so that there are two genders (masculine and feminine) instead of three (neuter), and the case system (by which grammatical relations in the sentence are indicated by various endings) also disappeared. They also share the fact that the adjective

typically comes after the noun it modifies, and it agrees with that noun in gender and number (singular or plural). The noun is often used with a definite article (*le, la, les* in French) and an indefinite article (*un, une, des* in French), both of which are placed before the noun. The informality/formality of situations is also characteristically reflected by the choice of pronouns of address, *tu/vous* 'you' in French.

During the course of time, French has lost or acquired features that did not exist in Latin and many of which are not found in the other Romance languages. Some of its most characteristic sounds are innovations: pharyngeal (guttural) R, nasal vowels as in *vin* 'wine' and *bon* 'good', and rounded front vowels as in *rue* 'street' or *peu* 'little'. It is true that adjectives agree with the nouns they modify in gender and number, but the gender of the noun is often not overtly given and simply has to be memorized, and for both adjectives and nouns the plural found in writing (with 's') is typically not sounded in the spoken language. Unlike Latin, French still has a rich verb morphology, but it has been greatly restructured since Latin times. The imperfect and simple past of Latin were augmented by a conditional and a new future tense, as well as a series of compound tenses (present perfect, pluperfect, future perfect, etc.) and other tenses with some sort of auxiliary verb (e.g. future with *aller* 'go'). French also has a subjunctive, although its use is often predictable: it is most often used in subordinate clauses after certain kinds of verbs in the main clause or after certain kinds of conjunctions. Modern French syntax is characterized by a fixed subject–verb–object word order; adjectives typically are placed after the noun they modify, as are relative clauses. Questions are asked by placing the subject after the verb, especially in formal usage, or by using the expression *est-ce que* 'is it that'. In the spoken language, questions are also often asked through a rising inflection (intonation).

Where French differs most is in how the verb is formed: the other Romance languages typically have suffixes that give the person (1st, 2nd, or 3rd) and number (singular, plural) of the subject of the verb, but French has lost most (but not all) of its verb suffixes. As a result, it has had to rely on nouns or pronouns (that come before the verb) to give that information; and typically, if no noun is present, then a pronoun has to be used. In most of the other Romance languages, the use of a subject pronoun is optional. Compare French *je chante* with Latin/Spanish/Italian *canto* 'I sing' (where *cant-* means sing and *-o* means 'I', and the use of the pronoun *ego/yo/io* is optional).

The vocabulary of French has also evolved over time. The lexical stock is essentially of Latin origin, but many learnèd words from Greek were added in the

Renaissance period, as were words borrowed from Latin. There are borrowings from numerous other languages (Germanic, Spanish, Arabic, Italian, and English), which reflect the history of contacts that the French people have had with other cultures.

Current Changes in Spoken French

Social background, sex, and age affect the way in which French is spoken. Young people from the outer suburbs of large French cities (many of whom are poor and/or from immigrant families) use more nonstandard features of the language, as well as *verlan*, a back-to-front slang, and borrowings from English and Arabic. The forms of language that are used are also influenced by the kind of linguistic situations in which speakers are involved. In the written language, the simple past is a marker of formal registers and in other usages is replaced by the present perfect form.

But what is perhaps most notable right now is the fact that the ordinary spoken language used by educated speakers is moving further and further away from the written language and the standard French spoken only 50 to 10 years ago. For example, subject pronouns like *je*, *tu*, and *il* are becoming prefixes, attached to the front of the verb, along with object pronouns like *me*, *te*, and *le*. The pronoun for 'we' is no longer *nous*, inherited from Latin, but *on*, which used to mean 'one'. The negative marker *ne* is also just about gone. In other words, twenty-first century French is evolving further and further away from its origins in Vulgar Latin and from the other Romance languages.

References

Ager, Dennis. 1990. Sociolinguistics and contemporary French. Cambridge: Cambridge University Press.
———. 1996. 'Francophonie' in the 1990s: problems and opportunities. Clevedon: Multilingual Matters.
Ball, Rodney. 1997. The French speaking world: a practical introduction to sociolinguistic issues. London: Routledge.
Harris, Martin. 1988. French. The Romance Languages, ed. by M. Harris, and N. Vincent. London: Routledge.
Lodge, Anthony. 1993. French: from dialect to standard. London: Routledge.
Offord, Malcolm. 1990. Varieties of contemporary French. London: Macmillan.
Posner, Rebecca. 1996. The romance languages. Cambridge: Cambridge University Press.
———. 1997. Linguistic change in French. Oxford: Clarendon Press.
Rickard, Peter. 1989. A history of the French language, 2nd edition. London: Hutchinson.
Walter, Henriette. 1994. French inside out, transl. by P. Fawcett. London: Routledge.

LINDA R. WAUGH

See also **Algeria; Indo-European 4: Romance**

Fromkin, Victoria Alexandra

Victoria Fromkin described herself as a 'professional radical revolutionary' in her youth (quoted in Cheng and Sybesma, 'Interview with Victoria Fromkin', *Glot International* 2:5.1(1996)), and the qualities of intellectual independence that sustained her in that work later came to characterize her career in linguistics.

She began the study of linguistics as a graduate student in 1961, when the American linguistic world was enmeshed in challenges posed by Noam Chomsky and generative grammar to the uniquely American form of structuralism developed in the 1940s and 1950s by followers of Leonard Bloomfield. Her first teachers at the University of California, Los Angeles (UCLA) were not part of the generative movement, but at an institute on machine translation in Venice, Italy, she met Yehoshua Bar-Hillel, a philosopher and colleague of Chomsky. Talking with Bar-Hillel, Fromkin's understanding of linguistics and the goals of linguistic theory were clarified. Chomsky's distinction between linguistic competence and linguistic performance and his conception of linguistic theory as an attempt to understand the nature of language and the human mind were considered revolutionary at the time, and that, of course, appealed to Fromkin.

Fromkin's doctoral dissertation was written under the direction of the linguist and phonetician Peter Ladefoged, and for some years thereafter she worked on the physical nature of speech sounds and the properties of a universal phonetic theory. Some linguists at the time rejected phonetics as part of the discipline, arguing that it dealt merely with physical phenomena, not with the structure of language. Fromkin's view of linguistics was more generous:

> The question of the boundaries of phonetics and linguistics, or whether such boundaries should be drawn, is an important one. . . . When I first came into the field I was interested in electromyographic registrations of linguistic units, and there were people who said: 'That is

not linguistics', and I said: 'But linguistics is whatever tells us more about the nature of human language and how language is realized in speech and in perception'. (*Phonetica* 37. 22–3, 1980)

It was just this ability to see beyond the constraints that some would impose upon the discipline that enabled Fromkin to expand her own work, and that of the field, in ways that led eventually to the cross-disciplinary specialization of neurolinguistics. For Fromkin, the research began with slips of the tongue. Her interest lay not in merely describing, categorizing, or even explaining slips and other speech errors, but rather in determining what such errors might imply about issues of language structure and language processing that were relevant to linguistic theory. In her landmark paper on speech errors published in 1971, she demonstrated that slips of the tongue involve discrete units of a language (for example, sounds, syllables, words) in accord with the rules of that language. That is, speech errors are not random. They are structured in the same way that language is structured, and studying such errors may provide insights into the way in which language is organized in the brain.

In her studies of lexical substitutions, for example (saying 'Take the milk out of the oven' when a speaker intended 'Take the milk out of the refrigerator'), Fromkin argued that such substitutions suggest that in our mental dictionary, words are specified with their semantic features. Other errors point to specification of phonetic, phonological, morphological, and syntactic features, as well as orthographic (spelling) properties. The vocabulary in our brain is not a mere alphabetical listing, but a complex set of independent and interconnected subsystems of linguistic information.

Fromkin's work on speech errors and their implications for linguistic theory was highly regarded and very influential. She edited two books of contributions by scholars from around the world, the first in 1973 (*Speech errors as linguistic evidence*), the second the result of a symposium she led at the 12th International Congress of Linguists in Vienna, Austria in 1977 (*Errors in linguistic performance: slips of the tongue, ear, pen, and hand*, 1980). In 1988, she contributed the state-of-the-art article on speech errors to *Linguistics: the Cambridge survey*. As her friends and colleagues were well aware, she always carried a notebook in which to record any speech errors she heard. The compilation eventually reached more than 12,000 examples.

In her later work, Fromkin investigated language use and loss in people with aphasia, dyslexia, and Parkinson's disease, and she promoted the study of sign languages and of language acquisition in children who had been denied language input by illness or neglect. In these studies, she explored evidence revealing the linguistic capacities of the human brain and the modularity of language.

Modularity—the theory that there is a unique language 'organ' distinct from other cognitive systems and abilities—was central to much of Fromkin's work, and her support for the theory of a genetically determined language faculty placed her within the general scope of Chomskyan generative linguistics. However, her work was often on the bridge between linguistics and other disciplines (aphasia studies, neurobiology, psychology), and so, except for debates over modularity, she was not a core figure in controversies of theoretical linguistics.

Partly because of this, and partly because of her strong administrative skills, she played a significant leadership role in national and international linguistic organizations, serving as president of the Linguistic Society of America (1985), as secretary and then chair of Linguistics Section Z of the American Association for the Advancement of Science (1993–1996, 1997–1998), and as an executive board member of the Permanent International Committee of Linguists. Her accomplishments were recognized by her election to the National Academy of Sciences.

In both her academic and her administrative work, Fromkin was always concerned with building bridges between linguistics and other disciplines: 'You can't do work on language without linguistics' (quoted by Cheng and Sybesma, 'Interview with Victoria Fromkin', *Glot International* 2:5.23 (1996)). She encouraged linguists to contribute their understanding of language to other fields, and she searched those fields for evidence that might support claims that linguists made about language and its organization in the brain.

Fromkin also took seriously the responsibility of bringing linguistics and its findings to the public and to the undergraduate population of colleges and universities. At UCLA, she taught the Introduction to Language class for more than 30 years, and it was from that class that she developed, with co-author Robert Rodman, her book *Introduction to language*, without doubt the most widely read of all introductory linguistics textbooks in the second half of the twentieth century.

Biography

Victoria Fromkin was born in Passaic, New Jersey, on May 16, 1923. She did her B.A. in economics from the University of California, Berkeley (1944). She married in 1948 and had one child. She returned to school in 1961, University of California, Los Angeles; and received her M. A. (1963) and Ph.D. with dissertation on phonetics guided by Peter Ladefoged (1965). She

was Assistant Professor, English, California State University, Los Angeles, spring 1965; joined the University of California, Los Angeles faculty in 1966 in Speech; in 1967 transferred to Linguistics and remained affiliated there throughout her career; in 1966 she was Assistant Professor, 1969 Associate Professor, 1972 Professor, Chair of Department of Linguistics 1970–1971 and 1973–1977, Dean of the Graduate Division 1979–1989, Vice-Chancellor of Graduate Programs 1980–1989, and she retired in 1991. She served on the faculty of the Linguistic Institutes of the Linguistic Society of America in 1966, 1976, 1977, and 1983; taught in the summer program in linguistics at the University of California, Santa Cruz 1971, 1972, 1973; and was visiting professor at University of Stockholm 1977, Cambridge University 1977, Wolfson College at Oxford University 1983 and 1987. She was President of the Linguistic Society of America 1985; President of Association of Graduate Schools 1988; Chair, Board of Governors of Academy of Aphasia 1991–1993; secretary 1993–1996 and then chair 1997–1998 of Linguistics Section Z of the American Association for the Advancement of Science; fellow of the Acoustical Society of America, the New York Academy of Sciences, the American Association for the Advancement of Science, and the American Psychological Society; executive board member of the Permanent International Committee of Linguists; elected member National Academy of Sciences, American Academy of Arts and Sciences, Aphasia Research Group of the World Federation of Neurology. She died in Los Angeles on January 19, 2000.

References

Cheng, Lisa, and Rint Sybesma. 1996. Interview with Victoria Fromkin. Glot International 2 (May). 1–22

Emmorey, Karen D., and Victoria A. Fromkin. 1988. The mental lexicon. Linguistics: the Cambridge survey, Vol. 3, ed. by Frederick J. Newmeyer, 137–8. Cambridge: Cambridge University Press.

Fodor, Jerry A. 1983. The modularity of mind. Cambridge, MA: MIT Press.

Fromkin, Victoria. 1971. The non-anomalous nature of anomalous utterances. Language 47. 27–62

———— (ed.) 1973. Speech errors as linguistic evidence. The Hague: Mouton.

————. (ed.) 1980. Errors in linguistic performance: slips of the tongue, ear, pen, and hand. New York: Academic Press.

————. 1980. Discussion [following 'The goal of phonetics, its unification and application', by Björn Lindblom]. Phonetica 37.

————. 1987. The lexicon: evidence from acquired dyslexia. Language 63. 1–22

————. 1988. Grammatical aspects of speech errors. Linguistics: the Cambridge survey, Vol. 2, ed. by Frederick J. Newmeyer 115–27. Cambridge: Cambridge University Press.

————. 1988. Sign languages—evidence for language universals and the linguistic capacity of the human brain. Sign Language Studies 59. 115–27.

————. 1991. Brain and language: redefining the goals and methodology of linguistics. The Chomskyan turn, ed. by Asa Kasher. Oxford, UK and Cambridge, MA: Basil Blackwell. 78–103.

————. 1993. Speech production. Psycholinguistics, ed. by Jean Berko Gleason and Nan Bernstein Ratner. Fort Worth, TX: Harcourt Brace Jovanovich.

————. 1998. Introduction to language, co-authored with Robert Rodman, 6th edition. Fort Worth, Tx: Harcourt Brace (previous editions 1974, 1978, 1983, 1987, 1992).

———— (ed.) 2000. Linguistics: an introduction to linguistic theory. Oxford, UK and Malden, MA: Blackwell Publishers.

Hyman, Larry M., and Charles N. Li (eds.) 1988. Language, speech and mind: studies in honour of Victoria A. Fromkin. London and New York: Routledge.

Jackendoff, Ray. 1994. Patterns in the mind: language and human nature. New York: Basic Books.

Pinker, Steven. 1994. The language instinct: how the mind creates language. New York: William Morrow.

————. 1999. Words and rules: the ingredients of language. New York: Basic Books.

Zwicky, Arnold M. 1980. Mistakes. Reynoldsburg, OH: Advocate Publishing.

JULIA S. FALK

See also **Bar-Hillel, Yehoshua; Chomsky, Noam; Neurolinguistics**

Function Words

Words are divided into two basic classes: (1) lexical, or open class words and (2) function, or closed class words. Lexical words refer to nouns (e.g. *dog, house*), verbs (e.g. *to go, to see*), and adjectives (e.g. *green, nice*), and supply the bulk of the meaning in a sentence. This class is called 'open' because speakers can freely add new words. Function words, on the other hand, as seen in these English examples, include determiners, such as *the* and *a(n);* auxiliary verbs (or simply 'auxiliaries'), such as *might, have,* and *be;* conjunctions, such as *and, that,* and *whether;* and degree adverbs, such as *very* and *too.* These words are

called 'functional' or 'grammatical' because they carry little meaning (have no synonyms) and typically 'help' another word. For instance, determiners add grammatical information about specificity and definiteness (*the dog* vs. *a dog*), but do not essentially alter the meaning of the phrase. They are also called 'closed class words', since speakers do not easily add new words to the set.

Lexical words typically carry intonational emphasis or stress, while function words are generally unstressed. Therefore, function words are prone to contraction—for example, the auxiliary *have* in *I've seen it*.

The distinction between function and lexical words has been very fruitful for linguistic description. So-called analytical languages, such as Chinese, use a large number of function words. In contrast, function words are typically lacking in the speech of young children, certain kinds of aphasia, and telegraphic speech. It is also widely confirmed that languages rarely borrow function words from other languages or invent new ones (hence their status as a 'closed' class). Most of the recent innovations in the English vocabulary, such as *pizza, angst, fax, E-mail, phat, AIDS, website, browser, screenager, to surf, Nethead,* and *technobabble,* are lexical rather than functional in nature.

Function words primarily add grammatical information, which means that they are defined above all by their syntactic behavior. Most traditional grammarians assumed (and their descendants continue to assume) that the structure of sentences and phrases is influenced mainly by lexical words. Function words were regarded as mere additions to lexical phrases. Thus, the sentence, '*The rabbit will see the fox*' was analyzed as a noun phrase *the rabbit*, followed by a verb phrase *will see the fox*. The determiner *the* was thus simply an addition to the noun phrase, and the auxiliary verb *will* was added to the verb phrase.

A shift in this thinking came in the 1980s within the framework of generative grammar. From then on, auxiliaries were seen as providing their own, independent, contribution to sentence structure. Nevertheless, function words still did not play a part in determining the category in which a phrase was placed—for example, a phrase such as *the rabbit* continued to be regarded as a noun phrase that happened to contain a determiner.

This view changed radically by the mid-1980s, when function words were increasingly treated as the key factor when deciding which sentence elements should be placed in which category. To use the technical terminology, function words were 'projecting to a phrase' or 'heading a phrase'. Determiners, for example, now came to be regarded as the head of determiner phrases—that is, *the rabbit* was now interpreted as a determiner phrase *the . . .* containing the noun phrase *rabbit,* instead of a noun phrase that happens to con-

tain a determiner. 'Functional projections' were thus assigned a structure and importance similar to 'lexical phrases'.

Research in the late 1980s and 1990s revolved around the question of exactly which functional projections a sentence may contain. Each function word that had a grammatical function was soon regarded as the main structural building block of a sentence. The increasing importance of function words in linguistic theories went hand in hand with an increasingly abstract description of sentence structure. This shift provided many empirical and theoretical advantages.

First, sentence structure could now be divided into three function domains: (1) a lexical domain around the verb, which establishes semantic relations between the main sentence elements; (2) a grammatical domain around the auxiliary, which establishes grammatical relations such as agreement (the auxiliary agrees in number and person with the subject: *I am/She is/They are leaving*); and (3) a discourse domain around the complementizer *that,* which links an embedded clause to a main clause (*I know **that** this is true* or *I wonder **whether** this is true*).

Second, differences between languages could be explained by how the function words, and the domains they define, were used. For instance, so-called verb-second languages, such as German, Middle English, Dutch, and Swedish, move the verb to the complementizer domain, whereas languages such as English refrain from doing so. For example, the word order in the English sentence, '*Yesterday the rabbit saw the fox*' would be, in German, '*Yesterday **saw** the rabbit the fox*'. Differences even between unrelated languages were thus reduced to very basic principles.

Function words and lexical words are not sharply distinct categories, but rather form a continuum. Certain classes of words can thus share features with both prototypical lexical words and prototypical function words. The English preposition is a case in point: some prepositions have lexical meaning, such as location (*behind*) and direction (*toward*), while others have little meaning (*of* or *to*). Many prepositions are used to introduce sentences (*after, for, like*) and are therefore similar to prototypical function words, namely, complementizers.

Grammatical meaning can be expressed in different ways. English uses independent auxiliaries to express present or past tense (*I am* thinking vs. *I was* thinking), but also inflects the verb for the same purpose (*I think* vs. *I thought*). Languages exhibit great variation along these lines: some languages express all grammatical meaning via independent function words, and such languages are called 'analytic'. In contrast, so-called 'synthetic' languages use inflection and other markings on lexical words throughout.

This distinction between analytic and synthetic languages also represents a continuum, and languages change in this respect over time. Old English made extensive use of grammatical markings on lexical words. Modern English has lost much of this capability and instead uses auxiliaries to fill the gap. In fact, the auxiliary verb *will* used to be a lexical verb in Old English, but its meaning ('to want') changed when it was recruited to express future tense. In modern theoretical approaches, such as the Minimalist Program, which tend to focus more on underlying differences than on surface variation, the distinction between analytic and synthetic languages becomes negligible.

Function words have little lexical meaning and usually carry no stress. In traditional grammars, they do not carry the function of governing an independent projection or phrase, whereas in some modern approaches they do. Function words are quite similar (and are historically related) to grammatical markers on lexical words.

References

Abney, Steven. 1987. The English noun phrase in its sentential aspect. Ph.D. Dissertation. Massachusetts Institute of Technology.

Chomsky, Noam. 1995. The minimalist program. Cambridge, MA: MIT Press.

Fukui, Naoki. 1986. A theory of category projection and its applications. Ph.D. Dissertation. Massachusetts Institute of Technology.

Gelderen, Elly van. 1993. The rise of functional categories. Amsterdam and Philadelphia: Benjamins.

Lightfoot, David. 1979. Principles of diachronic syntax. Cambridge and New York: Cambridge University Press.

Radford, Andrew. 1990. Syntactic theory and the acquisition of English syntax. Oxford and Cambridge, MA: Blackwell.

———. 1997. Syntactic theory and the structure of English. Cambridge and New York: Cambridge University Press.

Traugott, Elizabeth, and Berndt Heine (eds.) 1991. Approaches to grammaticalization, 2 vols. Amsterdam and Philadelphia: Benjamin.

ELLY VAN GELDEREN

See also **Auxiliaries; Grammar, Traditional**

Functional Approaches

Functional linguistics appeared as a reaction to formal approaches to grammar, especially generative and transformational approaches. One of its basic assumptions is that language is a symbolic system with a certain purpose or purposes, mainly communication, although there are other possibilities too, such as the use of language as an instrument of thought.

Apart from structures and form, any linguistic system also has functions. Functional approaches to language assume that there is a correspondence between form and function, and this correspondence is always motivated. However, a very subtle theoretical matter is what kind of function we are referring to. Apart from the most general functions of communication and organization of thought, it is possible to refer to functions at more atomic levels (i.e. functions of linguistic elements or functions of linguistic constituents, considered in different levels of linguistic analysis such as phonetics, morphology, semantics, syntax, etc.). At the most general level, however, there are several typologies with different functions that try to provide different functional alternatives, like the ones proposed by authors such as Karl Bühler (cognitive, expressive and conative/instrumental), M.A.K. Halliday (ideational, interpersonal and textual), or Roman Jakobson.

Today, there are two main approaches that claim to be 'functional'. These are M.A.K. Halliday's systemic functional linguistics, and Simon Dik's functional grammar, both discussed below.

Michael Halliday's systemic grammar follows the tradition of the London school and, more concretely, the work of John Rupert Firth, who developed his own theory studying language as part of a social system. According to Firth, language was used with a specific purpose in situational contexts, which, in turn, reflected cultural contexts. Other minor influences on Halliday's work are the linguistic theories of the Prague school around Nikolay Trubetskoy, Louis Hjelmslev's glossematics, and the ideas of the American linguist Benjamin Lee Whorf.

In 1961, Halliday, one of Firth's disciples, developed from Firth's ideas a theory that dealt with different scales and categories. He used the categories of structure, system, unit, and class, as well as three scales (rank, delicacy, and exponency) that connected the categories with one another and with the data. This grammar received criticism, especially concerning the categories of class and structure, the scale of rank, and the taxonomic nature of the theory, as mentioned by Christopher Butler (1985:29–38; 1995:529). Later, this grammar evolved toward what was first called

systemic grammar (Halliday 1967/1968), and some years later, systemic functional grammar (Halliday 1973, 1985), which is the current form of the theory.

Systemic grammar is based on a distinction between three main meta-functions of language: (1) the ideational function, which deals with the expression of content and with the experience of the speaker within the real world; here, it is possible to distinguish two subfunctions: experiential and logical; (2) the interpersonal function, which is used for establishing and maintaining social relations; and (3) the textual function, which deals with the creation of texts and the relations that are established within them. These meta-functions occur simultaneously in language. This simultaneity can be applied to two axes for the organization of the theory: the systemic (paradigmatic) and the structural (syntagmatic) axes. Each of these meta-functions involves different systems that can be organized according to the different units of the scale of rank (clause, phrase, group, word, informative unit). In fact, these three descriptive dimensions of the theory eventually become even more complicated, because Halliday also talks of a stratification in phonology, lexicogrammar, semantics, and context. The use of many axes for the organization of this theory makes systemic grammar a complicated one that tries to gather the complexity of language into many dimensions. In this respect, it does not have among its priorities the criteria of parsimony and elegance that appear in other theories. Additionally, systemic grammar is characterized by its being a theory that is sociologically oriented: that is, it classifies different contextual parameters of a social nature into dialectal and diatypical (based on different registers), and the latter into: field, which is the type of social activity in which language is inserted; tenor, which concerns the role relationships of roles among those who interact; and mode, which refers to the medium of communication. These types of parameters are related to the meta-functions of language: field to the ideational function, tenor to the interpersonal function, and mode to the textual function. Halliday's theory encompasses all usage domains of language, and in that respect it can be rightly considered a functional theory.

Simon Dik's Functional grammar presents a different approach. It claims to be a functional theory, since language is conceived mainly as an instrument of social interaction, and this characteristic is incorporated into the model. Despite this claim, the results, both in this first version of the theory (1978) and in the second (1989), are not as satisfactory as Halliday's proposal. Dik's grammar was influenced by the work of many previous authors, like Joseph Greenberg, James Fillmore, David Perlmutter and Paul Postal, Emmon Bach, H. Paul Grice, Herbert Clark, and Susan Haviland, among others. The architecture of its

model receives ideas already put forward by William Foley and Robert Van Valin in 1980 in their *Role and reference grammar*, but their proposal is original and different.

Dik (1989) proposes adequacy standards for his grammar, such as psychological adequacy, typological adequacy, and pragmatic adequacy, without which it would not be possible to account for language as an instrument of social interaction, and therefore the grammar would no longer be a functional grammar. However, the proposed model presents a high degree of formalization that is characterized by having an internal architecture with different levels that appear included one within another. The resulting embedding could be represented as: $[_{Level\,4}\,[_{Level\,3}\,[_{Level\,2}\,[_{Level\,1}$ Nuclear Predicate$]]]]$. Any clause has this basic configuration, to which the rules of expression can be applied in order to produce the concrete form of the sentence in English. In a posterior elaboration of the model by Kees Hengeveld (2004), the mentioned scheme constitutes the representational level, but this level is produced from information in the interpersonal level, and it is determined as much by the cognitive context as by a *cognitive component*.

In this new formulation of Dik's functional grammar, the roles of the interpersonal factor and the communicative context appear more clearly, which characterize this model as even more functional.

References

Butler, Christopher S. 1985. Systemic linguistics: theory and applications. London: Batsford.
———. 1995. Systemic functional grammar. In Verschueren, Östman, and Blommaert.
Dik, Simon C. 1978. Functional grammar. Amsterdam: North-Holland.
———. 1989. The theory of functional grammar, part I: the structure of the clause. Dordrecht: Foris.
Halliday, M.A.K. 1961. Categories of the theory of grammar. Word 17. 241–92.
———. 1967/1968. Notes on transitivity and theme in English. Parts 1, 2 and 3. Journal of Linguistics 3(1). 27–81, 3(2). 199–244, 4(2). 179–215.
———. 1973. Explorations in the functions of language. London: Edward Arnold.
———. 1985. An introduction to functional grammar. London: Edward Arnold, 2nd edition, 1994.
Hengeveld, Kees. 2004. The architecture of a functional discourse grammar. A new architecture for functional grammar, ed. by J.L. Mackenzie and Gómez González M.A. Berlin/New York: Mouton de Gruyter.
Nuyts, Jan. 1995. Functionalism vs. formalism. ed. by In Verschueren, Ostman and Blommaert.
Verschueren, J., J.-O. Ostman, and J. Blommaert (eds.) 1995. Handbook of pragmatics. Amsterdam and Philadelphia: John Benjamins.

CARLOS INCHAURRALDE

See also **Firth, John Rupert; Halliday, Michael Alexander Kirkwood**

G

Gender and Language

One meaning of the term *gender,* in the phrase 'Gender and language', is gender *as a characteristic of language as an abstract system.* In this sense, *gender* refers to a *grammatical* category. A second meaning of *gender* refers broadly to the 'human' and necessarily social concepts of femininity and masculinity.

Gender as Grammatical Category

Traditionally, languages have been described as having either 'natural' or 'grammatical' gender. A language with natural gender requires that the gender of an animate noun or pronoun corresponds to the biological sex of the person or animal to which that noun refers. In this way, *woman* and *girl* in present-day English are feminine nouns and *she* and *her* are feminine pronouns. Similarly, *bull* is a masculine noun and *vixen* is a feminine noun. Nouns like *computer* and *marmalade* are neither feminine nor masculine, but rather 'neuter'. Despite a very few odd but well-known exceptions (like a ship being sometimes referred to as *she*), the English language is usually seen as having natural gender.

Grammatical gender, in contrast, is 'formal'. In languages with grammatical gender, all nouns have a gender, although whether this is masculine, feminine, or neuter is unlikely to be evident from the noun itself. In French, for example, *chaise* (*chair*) is feminine and *pain* (*bread*) is masculine. The determiner (in these cases, the definite article) indicates the gender (*la chaise, le pain*).

Some masculine nouns and pronouns in English and other languages supposedly have the potential to be 'generic' or 'sex-indefinite', i.e. to have the capacity to include both males and females. Examples include *man* (and its compounds) and *he* (as well as some animal names such as *dog* and *lion*). Thus, in principle, it is grammatically correct to say to a class of male and female students: 'Everyone will get his homework back tomorrow.' But since *his* here is intended to include females, this illustrates that gender in English is in fact in part grammatical.

The genericity of those masculine forms in English that can in principle additionally refer to 'humans' has however been seriously challenged—one reason being that phrases like 'Man breastfeeds his young' sound odd; another that what may be intended as generic may not be so interpreted (by either sex), and accordingly that women and girls may be effectively excluded or rendered relatively invisible by the use of these 'generics'. The 'generics' *man* and *he* have thus been seen by feminists as one form of 'sexist language' (other forms being those that define, stereotype, trivialize, and/or degrade women, e.g. *Miss/Mrs* (*vis à vis Mr*), *usherette, air hostess, 'she's a blonde', dumb blonde*). (see Lakoff (1975) for an influential—although, in retrospect, problematic—pioneer account of such language.)

The masculine 'generics' as well as other 'sexist language' items now have a somewhat old-fashioned ring to them, and are often substituted for by alternatives (although not replacements) such as *people, he or she, s/he,* 'singular *they*' (especially in spoken English, e.g. 'Everyone will get their homework back tomorrow'), *flight attendant*, and *Ms*. These alternatives are now included in grammars and dictionaries, and

'inclusive language' is required by many journals and institutional Codes of Practice. Other nonsexist alternatives have been adopted for languages other than English (Pauwels 1998).

The vociferous and in many ways effective campaign against sexist language (see Lakoff (1975) for an influential—although, in retrospect, problematic—pioneer account of such language) was loosely based on a 'structuralist' view of language—one form, one meaning. Several limitations of this have now been identified, in particular that (a) word meanings change, (b) the meaning of a word will vary with context, (c) people will *interpret* a given word in a whole range of ways, and, perhaps most importantly, (d) sexist *discourse* can very easily occur without a single 'sexist language item' (for example, *people* can be used in a sexist way if it actually refers only to males; a text can still be degrading to women without using any 'sexist language'). Partly because of the evident limitations of 'nonsexist language', feminist critique has shifted to *discourse*, in the sense of 'ways of structuring knowledge and social practice' (Fairclough 1992), or of 'ways of seeing the world'. Gendered discourses *represent* men and women in particular ways, e.g. in contemporary childcare texts, 'Father as baby entertainer' (see Sunderland 2004). Other contemporary sexist discourses can be seen as (a) 'universal heterosexuality', i.e. the apparent assumption that every adult is either with a partner of the opposite sex or seeking one, (b) 'vive la diffrénce', i.e. that apparently essential differences between women and men are enjoyed by both and should/need not be problematized, (c) 'the battle of the sexes', i.e. that women's gains represent losses for men, (d) 'how to get your man and keep him' (Hollway 1984), a discourse *sine qua non* of many women's magazines, and (e) 'male sex drive' (Hollway 1984) (or 'men can't help their sexually aggressive behavior'). Masculinity and femininity, as well as individual women and men, and indeed gender relations, can be represented in particular ways in discourse in a whole range of written and spoken texts (Mills 1995). But representation can only ever be part of the story—it says nothing about how a text will be responded to by a reader.

Gender as Social Concept

The second, social use of the term *gender* in the context of gender and language study has witnessed considerable changes in meaning. Feminist linguistics in the 1970s largely bought into the traditional, sociolinguistic, variationist paradigm, in which gender was broadly mapped onto the biological category of *sex*. Although the feminist view of gender was always that it was primarily social or cultural rather than biological,

the overriding research question was whether and how women and men used language differently.

This project was characterized by two different approaches. The first was the 'dominance' paradigm, i.e. a concern with mixed-sex talk and whether men actually dominated women linguistically, making their use of language a form of 'doing power' (Kramarae 1981; Thorne et al. 1983). A British pioneer was Dale Spender, who in the 1980s popularized what were then new and important findings about male dominance in classrooms, in terms of boys being the recipients of most teacher talk, and the producers of most classroom talk (Spender 1982).

The other, slightly later approach was the more liberal '(cultural) difference' paradigm, exemplified in its most extreme form by Deborah Tannen (1990), in a less extreme form by Jennifer Coates (1998). Coates' work concerns particularly single-sex (women's) talk. In this approach, in which dominance was much less of an issue, men and women were seen as speaking differently as a result of having grown up in different linguistic subcultures.

However, these apparently different approaches also had a great deal in common. Focusing variously on women's disadvantages ('dominance') and women's strengths ('cultural difference') in their concerns with identification of gender *differences* in talk, they played down *similarities*, and underemphasized differences 'within' men as a group and women as a group. With exceptions, they both tended to generalize (women do this, men do that), to underplay the importance of context and local meanings, and to stress the importance of gender over other identities (see below). More profoundly, they implicitly shared a model of gender as broadly 'fixed', and somehow shaping or even *determining* talk, and accordingly linguistic gender differences—rather than of gender as itself being shaped *by* language use.

A combination of impatience with difference/dominance, and developments in cultural studies and feminist theory, led feminist linguists to turn the previous conception of the language/gender relationship on its head, i.e. from 'How does gender shape language?' (or 'How is language a reflection of gender'?) to 'How does language (or discourse) shape gender?' (especially over the last 20 years) Accordingly, one of the major ways in which the language and gender field (particularly that with a sociolinguistic orientation) has changed (especially over the last 10 or 15 years) is the growing recognition of the *problematic* nature of a model of gender as fixed (established at the age of 16 or 18 or whatever), and the newer understanding of gender as changing all the time.

Corresponding to this is a recognition that if gender is not fixed, the binary, 'gender differences' approach

to language use (which underpinned both the 'dominance' and the '(cultural) difference' paradigms) is also problematic (see Cameron (1992) for a useful critique). Gender and language study now acknowledges the further complexity of gender since it can be seen as coexisting with considerations of, inter alia, class, ethnicity, and sexuality.

One response to the question of *how* language shapes gender is 'through discourse, which is both shaped by *and shapes identity*'. To see gender (masculinity, femininity) as identity excludes neither shaping by social forces nor an individual's own agency (Bucholtz et al. 1999; Norton 2000). Neither does it entail fixity. Johnson for example sees masculinity and femininity as 'on-going social processes dependent upon systematic restatement' (1997:22).

If identity is further conceptualized as *multiple*, gender can be seen as one of several *identities* (e.g. masculinity/femininity, ethnicity, sexual identity)—although one identity may be foregrounded at any given time. Hence, while, say, ethnic identity might be experienced as most salient for some individuals in some contexts (for themselves and/or for others), this cannot be seen as separate from gender. Each identity will mediate and be mediated by the others.

Masculinity, already referred to in passing, is one recent concern of gender and language study (see Johnson and Meinhof 1997; Coates 2003). Masculinity can be manifested and constructed linguistically in a range of ways, as can femininity. Many of these are likely to be forms of what has been referred to as hegemonic (crucially, heterosexual) masculinity. The study of gender identities, both masculinities and femininities, from gay as well as the previous heterosexual perspectives has thus paved the way for 'Queer Theory', which rather than advocating inclusion and equal opportunities, entails problematizing all sexual identities, and thus advocates continual enquiry as regards heterosexual and other 'hegemonic' identities as well as more marginalized ones (Butler 1999; Nelson 1999).

Gender and language study does not however conceptualize gender only as identity. It can also be seen as *performance*—of doing rather than of being (Butler 1999). Butler does not deny identity—rather, she sees gender identity as coming from, rather than preexisting, performance. For Butler, people perform an identity—linguistically, in discourse, and in other ways. To illustrate this, looking at the talk of gay males, it is possible to distinguish between 'Gayspeak', the identifiable talk of gay men (a form of talk whose existence has been challenged)—and the idea of gay men deliberately speaking in a stereotypically camp way in order to signal their sexual identity. This 'performance' may however bear no relation to the way in which gay men speak when they are not interested in signaling their sexual identity.

Thus, it is now possible to see the previous 'two prongs' of gender and language research as having dovetailed into *discourse*, and discourse *as social practice* as a shaper of identity. The relationship between language/discourse and gender is now explored across a range of genres and settings (see e.g. Sunderland 2000, 2004). However, one particular recent and influential context-related concept for feminist linguistics is that of 'Community of Practice' (Holmes and Meyerhoff 1999). A 'Community of Practice' can be defined both by its membership *and* its practice. The 'practice', both nonlinguistic and linguistic, will include heterogeneity as well as homogeneity in the ways things are done, and the way they are talked about. It can be, and normally is, both gendered and gendering.

To conclude, gender and language study is now characterized by considerations of (i) gender as *identity*, and this identity as both multiple and fluid ('becoming' rather than 'being'), (ii) *representation*, in a range of written and spoken texts and text types, (iii) the ongoing social and linguistic *construction* of gender, (iv) individual agency, (v) the possibility of seeing gender as 'performance', (vi) discursive practices and discourse(s) as *social practice* (Fairclough 1992), and (vii) gendering through a 'Community of Practice' (see Bergvall et al. 1995; Hall and Bucholtz 1995; Bucholtz et al. 1999).

References

Bergvall, Victoria, Janet Bing, and Alice Freed (eds.) 1996. Rethinking language and gender research. London: Longman.

Bucholtz, Mary, A.C. Liang, and Laurel Sutton (eds.) 1999. Reinventing identities—the gendered self in discourse. New York: Oxford University Press.

Butler, Judith. 1999. Gender trouble: feminism and the subversion of identity, 2nd edition. New York: Routledge.

Cameron, Deborah. 1992. Feminism and linguistic theory, 2nd edition. London: Macmillan.

Coates, Jennifer. 1998. Women talk. Oxford: Blackwell.

Coates, Jennifer. 2003. Men talk: Stories in the making of masculinities. Oxford: Blackwell.

Fairclough, Norman. 1992. Discourse and social change. London: Polity Press.

Hall, Kira, and Mary Bucholz (eds.) 1995. Gender articulated: language and the socially constructed self. New York: Routledge.

Hollway, Wendy. 1984. Gender differences and the production of the subject. Changing the subject, ed. by J. Henriques et al. London: Methuen.

Holmes, Janet, and Miriam Meyerhoff (eds.) 1999. The community of practice: Theories and methodologies in language and gender research. Language in Society 28. 173–83.

Johnson, Sally, and Ulrike Meinhof (eds.) 1997. Language and masculinity. Oxford: Blackwell (includes 'Introduction' by Sally Johnson).

Kramarae, Cheris. 1981. Women and men speaking. Rowley, MA: Newbury House.

Lakoff, Robin. 1975. Language and woman's place. New York: Harper and Row.

Mills, Sara. 1995. Feminist stylistics. London: Routledge.

Nelson, Cynthia. 1999. Sexual identities in ESL: Queer theory and classroom enquiry. TESOL Quarterly 33(3). 371–91.

Norton, Bonny. 2000. Identity and language learning: gender, ethnicity and educational change. Harlow, Essex: Longman/Pearson Education.

Pauwels, Anne. 1998. Women changing language. London: Longman.

Spender, Dale. 1982. Invisible women: the schooling scandal. London: The Women's Press.

Sunderland, Jane. 2000. Review article: issues of language and gender in second and foreign language education. Language Teaching 33(4). 203–23.

Sunderland, Jane. 2004. Gendered discourses. London: Palgrave Macmillan.

Talbot, Mary. 1997. Language and gender: an introduction. Cambridge: Polity Press.

Tannen, Deborah. 1990. You just don't understand. New York: William Morrow.

Thorne, Barrie, Cheris Kramarae and Nancy Henley (eds.) 1983. Language, gender and society. Rowley, MA: Newbury House.

JANE SUNDERLAND

See also **Gender: Class Marking**

Gender: Class Marking

Nouns can be marked according to number, but there is also the possibility of marking nouns according to their belonging to certain classes. These classes were originally based on semantic criteria, which may still be present in many nouns, but it is customary to find nouns that do not match the criteria in their own class. In English and Indo-European languages in general, we normally refer to class marking as 'gender', and the original semantic basis for the marking is sex. Since sex applies to most animate beings, it is easy and straightforward to distinguish between male and female, which are marked grammatically as masculine and feminine, respectively. Inanimate beings, on the other hand, cannot be either male or female, and that is why in some Indo-European languages there is still a third class, the neuter, which marks grammatically those nouns denoting entities that have no sex. The picture becomes complicated by the fact that many nouns are classified within the different gender classes regardless of semantic considerations. Some examples follow.

In German, there are three genders, which reflect the semantic distinction in terms of sex in nouns like *Mann,* 'man', which is masculine, or *Frau,* 'woman', which is feminine. However, there are also cases in which there is no correlation with the semantic notion of sex, as can be seen in nouns like *Mädchen,* 'girl', which is neuter. In Spanish, there is no neuter, but there is a neat distinction between masculine and feminine, which closely follows the sex notion in nouns like *hombre,* 'man', and *mujer,* 'woman', but which is arbitrarily assigned in many other cases. In these two languages, gender is marked by special endings and the required agreement with certain forms of the article and demonstratives. Gender is also present as a grammatical distinction that requires different forms for personal pronouns. In English, there are no distinctive endings for nouns according to gender (although nouns sometimes have special forms that reflect gender, e.g. *heroine* is a feminine noun and *hero* is a masculine noun). In addition to this, there are no gender endings for articles and demonstratives. The only word class that should be clearly marked at all times according to gender is the class of personal pronouns, following these criteria: male humans (or animate beings in some situations) are masculine (*John came, but **he** left early*), female humans are feminine (*Mary didn't come. **She** stayed at home*), animals and plants in general are neuter (*I saw a cat yesterday. **It** was black*), and inanimate beings are neuter (*I like this bicycle because **it** does not weigh much*). There are some special cases that do not follow these rules (e.g. *Isn't **she** a beautiful BMW?* refers to a car, which is an inanimate being).

Class marking can be more complex than what gender alone suggests. In Bemba, a Bantu language, we find 16 different classes, which are marked morphologically in combination with number, corresponding to the original semantic values of humans, plants, fruits, mass/liquid, small objects or inanimates, animates, inanimate elongated objects, and paired body parts. The marking is carried out by means of prefixes (e.g. in *umu-ana*, 'child', the morpheme marking class-inclusion is *umu*). Swahili, another Bantu language,

has 11 classes, with the following prefixes for nouns, adjectives (which agree with the noun they accompany), and pronouns:

Class	Nominal Prefix	Adjectival Prefix	Pronominal Prefix
1	*m-*	*m-*	*a-/m-; yu-; w-; ye-*
2	*wa-*	*wa-*	*wa-*
3	*m-*	*m-*	*u-*
4	*mi-*	*mi-*	*i-*
5	*zero or ji-*	*zero or ji-*	*li-*
6	*ma-*	*ma-*	*ya-*
7	*ki-*	*ki-*	*ki-*
8	*vi-*	*vi-*	*vi-*
9	*zero or n-*	*zero or n-*	*i-*
10	*zero or n-*	*zero or n-*	*zi-*
11	*u-*	*m-*	*u-*

Swahili's classes can be grouped according to semantic criteria (similar to those of Bemba), but there can be many inconsistencies in the groupings. A classification proposal that takes into account all the nouns that do not fit into the system divides the noun classes into two subsets: a 'derived' set of classes, in which there is a fit between meaning and form (this is also called 'fixed' gender), and an 'inherent' set of classes, with arbitrary membership ('free' gender). There is overlap between these two sets formally, but the distinction helps in making the system consistent, because we can tell which nouns are classified according to semantic criteria and which are not.

As for the linguistic form used, Swahili and Bemba mark class membership by means of prefixes; but, as we have seen, there are other means. In Indo-European languages, such as Spanish, German, or Russian, class marking is in the ending. That is, classes are marked by means of information conveyed by suffixes. In Spanish, *-a* is normally the ending of feminine singular nouns, whereas *-o* is normally the ending of masculine singular nouns. However, this is not always so, because there are some nouns that end in *-a* which are masculine (e.g. *poeta*, the Spanish word for 'poet'). In this language, gender information is combined with number, so that all plurals add *-s* to the singular form, and we can find in general the following pattern:

	Masculine	Feminine
Singular	-o	-a
Plural	-os	-as

However, there are numerous exceptions to this rule, and there are also other endings, which are difficult to assign to one of the two classes. In German or

Russian, it is even more difficult to establish correspondences between suffixes and gender, since the endings also reflect grammatical case. The forms that can be used in German for the nouns Tisch 'table', Körper 'body', Mensch 'man' (masculine), and Jahr 'year', System 'system' (neuter) in the singular are the following:

	Masculine		Neuter	
Nominative	*-/-/-*	*(Tisch, Körper, Mensch)*	*-/-*	*(Jahr, System)*
Accusative	*-/-/-en*	*(Tisch, Körper, Menschen)*	*-/-*	*(Jahr, System)*
Genitive	*-es/-s/-en*	*(Tisches, Körpers, Menschen)*	*-es/-s*	*(Jahres, Systems)*
Dative	*-/-/-en*	*(Tisch, Körper, Menschen)*	*-e/-*	*(Jahre, System)*

The feminine and plural forms are simpler than these, since they are in most cases invariable; but it may be impossible to find out gender simply by looking at the ending (e.g. *Jahre* is the nominative plural and the dative singular of *Jahr,* 'year', which is neuter, whereas *Frage,* 'question', is the nominative singular of a feminine word; all these forms end in *-e*).

Articles and adjectives normally agree in gender with nouns in Indo-European languages (English is one of the exceptions), which means that they have full-ending paradigms for masculine, feminine, and, in some cases, neuter. Pronouns also use different ending paradigms according to gender, so that they can agree with the gender and number of the noun that corresponds to the entity to which they refer.

References

Corbett, Greville. 1991. Gender. Cambridge: Cambridge University Press.

Craig, Colette (ed.) 1986. Noun classes and categorization. Amsterdam and Philadelphia: John Benjamins.

Givón, Talmy. 1983. Syntax. A functional–typological introduction. Amsterdam and Philadelphia: John Benjamins.

Heine, Bernd. 1982. African noun class systems. Apprehension. Das sprachliche Erfassen von Gegenstanden, Vol. 1, ed. by H. Seiler and C. Lehmann. Tübingen: Narr.

Heine, Bernd, Ulrike Claudi, and Friederike Huennemeyer. 1991. Grammaticalization: a conceptual framework. Chicago: University of Chicago Press.

Matthews, Peter H. 1991. Morphology. Cambridge: Cambridge University Press.

CARLOS INCHAURRALDE

See also **Gender and Language**

Generation

Generation is the process by which thought is rendered into language. Within computational linguistics, it is referred to as 'natural language generation' (NLG) to help distinguish it from Chomsky's generative grammar and to contrast with natural language understanding. It is the study of how actual speakers, people or computers, construct utterances in actual contexts—the situations that motivate them to speak. As such, NLG is part of the larger field of Cognitive Science, where it is also referred to as 'production', particularly by psycholinguists.

The first NLG systems were developed in the 1950s as part of machine translation systems. The field gained maturity in the 1980s, with its own series of workshops and conferences and its own unique problems. As gauged by membership in the special interest group SIGGEN, there are several hundred people today actively pursuing research on generation.

NLG takes its methodology from Artificial Intelligence. To study a cognitive capability, you design and implement computer programs that attempt to replicate it: in this instance to produce fluent utterances for a purpose. To do this, the program ('generator') starts with a body of data or information. This might be the daily movement of a stock market index (Kukich 1988) or numerical data about temperature and wind patterns (Goldberg et al. 1994). In cases like these, the first thing the generator must do is analyze the data to determine both what information it contains (e.g. what is particularly salient: have winds increased or diminished) and what concepts—ultimately what words or phrases—could be used to communicate that information.

There is a consensus among NLG researchers that generation involves three broad-brush components: (1) determining and organizing the information content to be expressed; (2) 'microplanning', where the sentential and referential structures are determined; and (3) 'surface realization', where the text plan is processed by a grammar to construct the sequence of syntactically and morphologically appropriate word forms, which are then rendered on some output medium, typically formatted text displays or web pages. (For details of alternative NLG architectures, see McDonald (2000) or Reiter and Dale (2000).)

Surface realization is the most advanced of these three components since it draws on the well-established knowledge of grammar in linguistics and computational linguistics as a whole. Virtually every kind of grammar that linguists have developed has been applied to NLG including some that are relatively unknown in the wider community such as Melcuk's Meaning-Text Theory and Halliday's Systemic Functional Grammar (SFL). Interest in these theories of grammar stems from the need to reason about the alternative choices that are available, the functions they perform, and their consequences for other choices later on. For example, saying '*the house is red*' vs. '*the red house*' will express the same fact about the house, but with differences in emphasis and in what parts of the sentence they leave open for other information to fill. Surface realization is also the only component within the architecture of NLG systems that is sufficiently mature for 'plug and play' reusable components to have emerged, notably KPML (Bateman 1997) and FUF/SURGE (Elhadad and Robin 2001) both of which use SFL.

The state of the art in NLG is measured by a combination of the fluency of the texts it is possible to produce and comparative difficulty of adapting to new subjects or genres. Consider this example, an automatically generated recipe for butter bean soup (Dale 1992: 14).

> Soak, drain and rinse the butter beans. Peel and chop the onion. Peel and chop the potato. Scrape and chop the carrots. Slice the celery. Melt the butter. Add the vegetables. Sauté them. Add the butter beans, the stock and the milk. Simmer. Liquidize the soup. Stir in the cream. Add the seasonings. Reheat.

Notice how this text is tailored to its genre. All the sentences are imperatives; objects can be omitted when they are obvious ('*Reheat __*'), and the sentences are simple and short. One of the problems in microplanning is how to formulate this 'tactical' knowledge about a genre's preferred constructions in such a way that it can be deployed by other NLG systems producing texts about a different subject, especially when they use different processing methodologies. We know how to do this with grammars, but not with the knowledge of how to balance the consequences of alternatives when a text is composed.

This class of problems, unique to generation, is further illustrated with the example below, which was produced by Robin's STREAK system (1993). It is an example of the best that can be done today as it is indistinguishable from what a human sports journalist would produce as a capsule summary of a basketball game.

Dallas, TX—Charles Barkley matched his season record with 42 points Friday night as the Phoenix Suns routed the Dallas Mavericks 123–97.

STREAK uses an architecture based on revisions to an initial draft, where it continually looks for opportunities to incorporate historical knowledge into a skeleton of reported facts. In this instance, for example, there is the fact that the Mavericks are on a long losing streak. It is not possible to add the number of losses to the sentence, given its present structure; however, this generator has extensive tactical knowledge about the choices available to it and knows that if it uses an alternative way of phrasing the fact that the Suns lost, one that reifies the loss as a noun, it can then incorporate the number of losses by modifying the noun with the count. Underlines indicate the text that has changed:

. . . the Phoenix Suns <u>handed</u> the Dallas Mavericks <u>their 27th defeat in a row at home</u> 123–97.

The focus of ongoing research is in text planning problems such as this illustrates; in the extension of established capabilities to larger texts (as this is written, the limit is multiple page, individually tailored instruction, or advice pamphlets), and in integration with other modalities such as real-time graphics and the production of speech with appropriate prosodics.

References

Bateman, John. 1997. Enabling technology for multilingual natural language generation. Natural Language Engineering 3. 15–35.

Dale, Robert. 1992. Generating referring expressions. Cambridge, MA: MIT Press.

Elhadad, Michael, and Jacques Robin. 2001. SURGE: a comprehensive plug-in syntactic realization component for text generation. Computational Linguistics, in press.

Goldberg, Eli, Norbert Driedger, and Richard Kittredge. 1994. Using natural language processing to produce weather forecasts. IEEE Expert 9(2).

Kukich, Karen. 1988. Fluency in natural language reports. Natural language generation systems, ed. by David McDonald and Leonard Bolc. New York: Springer-Verlag.

Levelt, Willem J.M. 1989. Speaking: from intention to articulation. Cambridge, MA: MIT Press.

McDonald, David. 2000. Natural language generation. Handbook of natural language processing, ed. by Robert Dale, Herman Moisi and Harold Somers. New York: Marcel Decker.

Reiter, Ehud, and Robert Dale. 2000. Building natural language generation systems. Cambridge, UK: Cambridge University Press.

Robin, Jacques. 1993. A revision-based generation architecture for reporting facts in their historical context. New concepts in natural language generation: planning, realization and systems, ed. by Helmut Horacek and Michael Zock. London: Pinter.

SIGGEN see www.aclweb.org

DAVID MCDONALD

See also **Computational Linguistics; Machine Translation**

Generative Grammar

Generative grammar is a conceptual model whose central tenet is that language is a property for which human beings are biologically prewired. While all models of language assume some role for both biology and environment, they differ with respect to the emphasis that they place on each: empiricist models attribute a greater role to the environment, and focus on differences across speakers' grammars (for instance, how the particular input that children receive influences the development of their grammar). Nativist models, in contrast, attribute a greater role to the biological component, and focus on commonalities across speakers' grammars. While all nativist models assume language to be biologically determined, they differ with respect to the nature of such knowledge: The emergentist approach, favored largely by psychol-ogists, attributes knowledge of language to general-purpose learning mechanisms, and thus assumes this ability to be no different from any other cognitive capability; the generative approach, in contrast, subscribed to by many linguists, views linguistic knowledge as being unique and specialized, and stemming from brain structures devoted specifically to the processing of language.

The generative framework had its origins in the 1950s with the publication of Noam Chomsky's 1957 book *Syntactic structures,* which built on the work of his teacher Zellig Harris. Chomsky's approach was a reaction to the behaviorist theory of language prevalent at the time, championed by the psychologist Skinner. Under a behaviorist model, the brain is considered a blank slate with regard to linguistic knowledge;

children must thus be explicitly taught their language by the adults around them in a stimulus–response manner, their behavior being rewarded when they imitate the adults' language correctly. Chomsky instead advocated a view subscribed to in the previous century (to which behaviorism had been a reaction) that some brain activities are unconscious and reflexive, just as is the case for many physical processes. Much of human beings' linguistic knowledge, Chomsky argued, is abstract and unconscious, but can be brought to conscious awareness by examining speakers' usage of such linguistic knowledge (known as the competence/performance dichotomy).

Under a generative approach, human beings are assumed to be prewired for language, beginning life not with a blank slate but rather with a linguistic template or blueprint that they flesh out upon exposure to specific linguistic data. Instead of learning language by imitating those around them, children create their own grammars. One reason for assuming this is that they make errors that adults do not (e.g. 'I hurted myself') and generate novel forms, neither of which should occur if they learned solely by imitating. In addition, most children are not given explicit instruction in their language or corrected on errors, necessary in an imitative model, yet all manage to acquire language. Furthermore, although the environments in which children acquire their language vary, they all go through similar stages in acquiring a language and during the same general time frame (for instance, forming sentences with content words such as nouns and verbs around 18 to 24 months, and function words like 'will' and 'my' between 24 and 30 months). Moreover, although children do not hear examples of every possible structural pattern, they nonetheless attain a grammar capable of generating all the possible sentences in their language (known as the poverty of the stimulus argument). And, although each is exposed to different data and in a different order, they all end up with the same basic grammar for their language, which would be unexpected under an imitative account.

The name 'generative grammar' is used to refer to this model since speakers are assumed to possess a grammar capable of generating all the possible sentences in their language (while excluding all the impossible ones). The grammar consists of a finite number of rules, yet is capable of generating an infinite number of sentences from such rules due to their ability to refer back to each other repeatedly (known as recursion). For instance, one can continue to embed sentences within one another as in the example 'John thought that Mary said that Fred believed that Cindy suspected that the student had read the book,' each further embedding resulting in a new sentence. Another central feature of such a grammar is that it is highly constrained. For instance, all syntactic rules make reference to the internal structure of the sentence (known as structure dependency). Thus, yes–no questions in English are formed by moving the auxiliary to the front of the sentence ('Will the student read the book?'), such a rule being framed in terms of an internal grammatical unit (the auxiliary 'will'). No language has rules that are structure independent, such as moving the third word to the front of the sentence to form a question, which refer instead to surface properties such as linear position. That the latter formulation will not work can be seen when the subject is replaced with a pronoun: while the structure-dependent rule will generate the correct question since it always moves the auxiliary ('Will he read the book?'), the structure-independent formulation will not, as the auxiliary is no longer the third word ('Read he will the book?'). Constraints such as structure dependency support the idea of speakers possessing unconscious, abstract linguistic knowledge, as there is no overt evidence of the syntactic groupings of words in the input that they hear. And, the fact that speakers were never taught such constraints, let alone being aware that they exist, coupled with the fact that these hold across all languages, supports the idea of there being a wired-in universal component to language.

While the idea of an innate, prewired blueprint has remained constant in generative grammar, the conceptual details have varied across the decades. In the 1960s, the emphasis was on the distinction between a deep structure, which conveyed the semantic properties of a sentence, and a surface structure, which supplied its pronunciation. Thus, the passive sentence 'The book was read by the student' was assumed to come from the same deep or underlying structure as its active counterpart, 'The student read the book', as both have the same meaning. In the 1970s, the emphasis shifted to finding the set of transformations used to derive the various syntactic patterns of each language. The list included transformations for passives, yes–no questions, and wh-questions ('Which book did the student read?'). While the number of sentences in a language is potentially infinite, it was assumed that the number of transformations could be reduced to a finite set. However, it soon became clear that there were many more transformations than it was possible to enumerate. Emphasis then shifted to narrowing down the transformations by type. Two general types were established: noun phrase (NP) movement, which moved a phrase within a sentence, as in passives, and wh-movement, which moved a phrase outside a sentence to

a presentential landing site, as in wh-questions. Eventually, these two transformations were collapsed into one general transformation, move alpha, which allowed movement of any constituent anywhere, subject to certain constraints.

In the 1980s, the model was flipped on its head. Rather than looking for the possible structures in a language, the emphasis instead shifted to determining the impossible structures. This represented a significant evolution in the conceptual model: whereas the number of patterns possible in a language is potentially infinite, the number of constraints is thought to be very small. In addition, it also made it possible to shift the emphasis to universal aspects of language, rather than simply to those properties that an individual language possessed. The grammar was now taken to consist, not of a set of rules, but rather of a set of autonomous modules that interacted with each other; one conception of such a model became known as government-binding theory, named after two of the modules, while alternative models were also proposed such as lexical-functional grammar. The overall conceptual model became known as the principles-and-parameters model since it considered language to consist of a set of wired-in principles that all languages shared, along with a set of parameters that they also shared, but whose values varied cross-linguistically and needed to be set upon exposure to language-particular data (such innate knowledge being referred to as Universal Grammar). An example of a principle would be a movement constraint known as subjacency, which prohibits movement of a phrase out of more than one clause or noun phrase in a single step (the name 'subjacency' referring to the fact that movement can occur to an adjacent clause, but not a subadjacent one). Thus, one cannot say 'Which book do you know the student who read?' since 'which book' has been moved from within the relative clause ('the student who read which book') and the upper sentence ('you know the student who read which book'). While such a constraint is thought to be universal, the constituents out of which the element may move (the bounding nodes) vary cross-linguistically. English is freer in its movement allowances than Russian, but less so than Italian or Swedish; a language like Japanese, in contrast, allows no overt syntactic movement. Thus, a parametric difference linked to this principle would dictate what the bounding nodes for a given language are.

In the 1990s, the emphasis turned to making the model even simpler conceptually. The new approach, called minimalism, assumed a much more limited role for the syntactic component. It now was seen as a computational device that simply checked that sentences were formed correctly. All morphology was assumed to be attached directly in the lexicon; the syntactic component then checked to see that features on the words matched. If so, the derivation was said to converge, otherwise, it crashed. Thus, the sentence 'The student enjoys the book' would be acceptable since 'student' and 'enjoys' are both third-person singular, whereas 'The student enjoy the book' would be ungrammatical as the number agreement on the noun and verb do not match. The formation of sentences was now assumed to occur by means of a few basic operations such as merge, used to generate basic declarative sentences, and move, used to derive patterns such as passives and questions. Another conceptual model that developed during this decade was optimality theory, which attributed variation among languages to their different rankings of a set of universal constraints.

Since any model proposed for a grammar must be one that is learnable by the child, language acquisition research has helped to shape development of the generative framework, and much first- and second-language research today is solidly grounded in such a framework. First-language researchers are interested in questions such as whether the principles of language are all present at birth or instead come online gradually as the child develops cognitively (the continuity/ maturation debate). Second-language researchers are interested in determining whether second-language learners have full access to the language faculty as first-language learners do, partial access only (properties that are the same transferring, but new ones not being acquirable), or no access at all. And, a central question for both first- and second-language acquisition is whether there is a critical time period during which language must be acquired, as is true for other biologically determined properties.

While the conceptual details of the generative grammar model have changed greatly over half a century, the basic underlying tenet, that language is a species-specific property for which human beings come prewired, has remained constant. Future research will undoubtedly yield new insights into the specific shape of the grammar, while remaining true to the model's belief in a wired-in blueprint.

References

Baker, Mark. 2001. The atoms of language: the mind's hidden rules of grammar. New York: Basic Books; Oxford: Oxford University Press, 2002.

Chomsky, Noam. 1957. Syntactic structures. Mouton: The Hague.

Chomsky, Noam, Adriana Belletti, and Luigi Rizzi (eds.) 2002. On nature and language. Cambridge and New York: Cambridge University Press.

Radford, Andrew. 1997. Syntactic theory and the structure of English: a minimalist approach. Cambridge and New York: Cambridge University Press.

Radford, Andrew. 1988. Transformational grammar: a first course. Cambridge and New York: Cambridge University Press.

Smith, Neil, and Deirdre Wilson. 1979. Modern linguistics: the results of Chomsky's revolution. Brighton: Harvester, and Indiana: Indiana University Press.

NAOMI BOLOTIN

See also **Chomsky, Noam; Harris, Zellig Sabbetai**

Genericity

Much of our knowledge about the world is expressed in sentences such as

(1) *A raven is black.*
(2) *Birds lay eggs.*
(3) *The tiger has stripes.*
(4) *Mary smokes.*

These are examples of generics.

Generics can occur in many syntactic forms: (1) involves the indefinite singular noun phrase *a raven*, (2) contains the bare plural noun *birds*, and the subject of (3) is the definite singular noun phrase *the tiger*, whereas that of (4) is the name *Mary*. All these forms can have nongeneric uses as well. Indeed, there does not seem to be any sentence *form* that is unique to generics; what characterizes them is their *meaning*. Intuitively, the above sentences seem to express some sort of generalization, about ravens, birds, tigers, and Mary. Yet, it is far from clear exactly what generics mean. What does it mean to say that some generalization holds?

Possibly the first hypothesis that comes to mind is that generics express some sort of quantification. *A raven is black*, for example, could be taken to mean that a certain number of ravens are black. But how many? What is the generic quantifier? It cannot be *every,* because *A raven is black* is true despite the existence of albino ravens. It cannot be *most*, since *Birds lay eggs* is true despite the fact that fewer than half of all birds lay eggs (adult females only). The problem is even harder with *Mary smokes*, because it is quite unclear how often Mary has to smoke in order for the sentence to be true.

Faced with the diversity of interpretations of generics, researchers have taken two general approaches. The first attempts to define the quantifier in such a way that its nature, possibly in conjunction with context, intonation, and world knowledge, may account for all the interpretations of generics. For example, one could still maintain that the generic quantifier is *most*, and propose that to determine the truth or falsity of *Birds lay eggs*, only female birds need to be considered. Since most female birds lay eggs, the sentence is true.

An alternative approach denies that there is any quantification involved. Rather, generics are evaluated with respect to rules and regulations. For example, *A raven is black* is true not because of the properties of individual ravens, but because there is a (biological) rule stating that ravens are black.

Perhaps one of the difficulties in determining whether generics are quantificational is that generics are *lawlike*; they cannot express accidental generalizations. For example, even if it turned out that all Supreme Court judges had an even Social Security number, the generic *Supreme Court judges have an even Social Security number* would be odd. Contrast this with *Supreme Court judges are appointed by the President*, which is fine. Intuitively, this is because there is a law stating that Supreme Court judges must be appointed by the President, but no law governs their Social Security number.

To account for lawlikeness, some researchers treat it as a form of necessity. According to standard accounts of necessity, a necessary statement is not merely true in the actual world, but in all *possible worlds*. That is to say, there are any number of ways the world could have been different from the way it actually is; but in all such hypothetical situations, a necessary statement would still be true. The idea is, then, that a lawlike statement is true in 'sufficiently many' possible worlds.

Indeed, it seems that the truth or falsity of a generic does not depend on specific events in the actual world. In the sentence *A computer computes the daily weather forecast*, for example, *the daily weather forecast* cannot be taken to refer to the current weather forecast. For example, if we know that today's forecast predicts a blizzard and is the main news item, we still cannot change the above sentence to *A computer computes the main news item*. Yet, if *the daily weather forecast* did refer to the current weather forecast, this last statement should be true.

However, although generics do not seem to depend on specific events in the actual world, their interpretation nevertheless depends on real-world circumstances.

Suppose the weather report is Mary's favorite newspaper column. Although the world could well have been such that Mary would have no interest in the weather forecast, we may still truthfully say that *A computer computes Mary's favorite newspaper column.*

Thus, the truth or falsity of the generic statement *A computer computes the daily weather forecast* depends on the actual world, rather than what may be the case in hypothetical circumstances, but the interpretation apparently cannot be tied to a specific point in time.

To account for this aspect of genericity, some theories propose to restrict the possible worlds applicable to interpretation only to those that are *normal*, or those that are close to the actual world in terms of its essential properties. Such theories have to face the problem of defining normality or essence in such a way that a world in which, say, Mary is not interested in the weather is not included in this definition.

Alternatively, it has been proposed that generics are evaluated with respect to a variety of possible futures (e.g. that tomorrow the weather forecast may not be the main news item) but not with respect to alternatives to the world as it actually is (e.g. that Mary might not ever have been interested in the weather forecast).

It is probably evident from this discussion that, despite much progress, generics remain a puzzling and deeply contested phenomenon, with many questions and few widely accepted answers. In fact, one of the interesting and, so far, unanswered questions is this: if it is so difficult to define what generics actually mean, why do we use them so often?

References

Carlson, Greg. 1977. Reference to kind in English, Ph.D. Dissertation. University of Massachusetts; New York: Garland, 1980.

Carlson, Greg. 1989. On the semantic composition of English generic sentences. Properties, types and meaning, ed. by Gennaro Chierchia, Barbara H. Partee and Raymond Turner, 167–92. Dordrecht: Kluwer.

Carlson, Greg, and Francis J. Pelletier. 1995. The generic book. Chicago: University of Chicago Press.

Cohen, Ariel. 1996. Think generic: the meaning and use of generic sentences. Ph.D. Dissertation. Carnegie Mellon University; Stanford: Center for the Study of Language and Information, 1999.

Cohen, Ariel. 1999. Generics, frequency adverbs, and probability. Linguistics and Philosophy 22. 221–53.

Dahl, Östen. 1975. On generics. Formal semantics of natural language, ed. by Edward L. Keenan, 99–111. Cambridge: Cambridge University Press.

Delgrande, James P. 1987. A first-order conditional logic for prototypical properties. Artificial Intelligence 33. 105–30.

Heyer, Gerhard. 1990. Semantics and knowledge representation in the analysis of generic descriptions. Journal of Semantics 7. 93–110

Pelletier, Francis J., and Nicholas Asher. 1997. Generics and defaults. Handbook of logic and language, ed. by Johan van Benthem and Alice ter Meulen, 1125–77. Amsterdam: Elsevier.

Schubert, Lenhart K., and Francis J. Pelletier. 1987. Problems in the representation of the logical form of generics, plurals, and mass nouns. New directions in semantics, ed. by Ernest LePore, 385–451. London: Academic Press.

———. 1989. Generically speaking, or using discourse representation theory to interpret generics. Properties, types and meaning, ed. by Gennaro Chierchia, Barbara H. Partee and Raymond Turner, 193–268. Dordrecht: Kluwer.

ARIEL COHEN

Genetic Relationship

If we look at the word for 'brother' in a handful of European languages, we can see some striking resemblances: for example, between English *brother*, German *Bruder*, Polish *brat*, Welsh *brawd*, Irish *brathair*, Swedish *broder*, Dutch *broer*, and Czech *bratr.* Clearly, there is some kind of relationship between these languages, but how exactly do we account for these similarities? And why is it that differences also exist?

The most salient way in which languages come to be related is through a genetic relationship. This means that, at some point in their history, the several languages in question were actually the same language (normally known as the protolanguage) and underwent a period of common development from which their similarities arise. This is the case, for example, with the major modern Romance languages (i.e. Italian, Spanish, Portuguese, French, and Romanian), which all have their origins in the Latin language spoken throughout the ancient Roman empire. As time progresses, however, divergence occurs, mostly as a result of population migrations or through loss of a central unifying standard: certain groups of speakers of the protolanguage are split off from one another and, subsequently, their now separated language varieties pursue their own courses of development, from which the differences between them arise.

This model of language relationship is usually known as the family tree model, and was proposed by August Schleicher in 1871. However, in its purest form, the family tree model gives rise to a number of problems. Firstly, it assumes that there are clear dividing lines between languages (both temporally and geographically), so that there is a point at which we can make a distinction between, say, late vulgar Latin as a protolanguage and early Spanish as a separate development from it; however, it is very difficult to draw these lines clearly. Secondly, the pure family tree model assumes that no further interaction takes place between the separated languages and the protolanguage after a split has occurred; however, this is again often not so, as Romance demonstrates—the protolanguage (Latin) continued to be used quite extensively alongside the modern Romance languages, for example, by the church and the universities. Similarly, the model also does not allow for a later convergence between languages that resulted from an earlier split. For example, it is estimated that roughly a third of the English vocabulary is of French origin. However, the ancestors of English and French split off from one another at a very early date, and this shared vocabulary does not stem from that more distant genetic relationship between English and French; rather, it is a result of large-scale later borrowing into English, especially in the years following the Norman conquest of England in 1066 CE. This example is salutary, because it shows that a large degree of similarity (e.g. in vocabulary) does not necessarily result directly from a genetic relationship.

Clearly, therefore, although genetic relationship is often a very important determining factor in language similarity, as with the Romance languages, the purest form of the family tree model must be considered an idealization of reality.

There are, in fact, broadly three other reasons (apart from genetic heritage) that can lead to a close interrelationship between languages. Probably the next most important, which has already been alluded to, is borrowing from one language into another. Although small numbers of loans are very common, large-scale borrowing that leads to a substantial overlap in vocabulary and/or grammar normally arises either in a context where many speakers are bilingual or in a situation where two languages are in close geographical proximity to one another. To account for this, another model of language relation was proposed by Johannes Schmidt in 1872. This is known as the wave model and it is based on the principles of dialectology. The wave model rejects the notion that precise geographical borders exist between different languages. Instead, it assumes that the spatial extent of individual linguistic features (be they sound patterns,

words, or grammatical features) is marked by equally individual borders (known as isoglosses). These isoglosses may well form relatively strong borders between languages when they bundle closely together, but they are by no means necessarily located in the same places. The wave theory proposes that individual features spread from one place to another by means of diffusion. In other words, the extent of a feature's isogloss spreads as it gradually moves from one location to the next, rather like passing a message along a chain of people; and, as more features spread out, the more similar the language spoken in the different places along the chain becomes. Because it does not assume that shared features are always the result of a single period of shared historical development, this model is able to accommodate later convergence between languages. Similarly, it also allows for the influence of non-genetically-related languages in these contexts, for example, the influence of the Semitic Arabic language on the Indo-European Urdu language.

The two other causes of language similarity are (1) chance similarity and (2) the operation of universal linguistic processes, which lead to similar developments after languages have split off from one another. As an example of the latter, we can consider the development of the definite articles (= 'the') in two of the Romance languages—French and Italian. These are descended in both languages from the Latin pronoun *ille* and its variants. In both French and Italian, the feminine singular forms are the same (*la*) and both also lose their /a/ vowel before another vowel (French *l'allumette*; Italian *l'amica*). However, this does not mean that the word *la* came either directly from Latin or was borrowed from Italian into French (or vice versa); rather, universal phonetic principles (loss of an unstressed vowel and the merging of adjacent vowel sounds) have operated on the original Latin word (*illa*) with the same end results.

References

Aitchison, Jean. 1991. Language change: progress or decay? Cambridge: Cambridge University Press.

Anttila, Raimo. 1972. An introduction to historical and comparative linguistics. New York: Macmillan.

Hock, Hans Henrich. 1986. Principles of historical linguistics. Berlin: Mouton de Gruyter.

Hock, Hans Henrich, and Brian D. Joseph. 1996. Language history, language change, and language relationship: an introduction to historical and comparative linguistics. Berlin: Mouton de Gruyter.

Jeffers, Robert J., and Ilse Lehiste. 1979. Principles and methods for historical linguistics. Cambridge, MA: MIT Press.

Lass, Roger. 1997. Historical linguistics and language change. Cambridge: Cambridge University Press.

ANDREW WILSON

See also **Indo-European 4: Romance**

Genre

'Genre' stems from Latin *genus* and French *genre*, words meaning 'kind'. Hence, the popular meaning of genre in *Do you have some shoes in the same genre*? is close to its etymological roots. Historically, however, and largely as an inheritance from Aristotle, 'genre' was defined normatively as a given category of literature, with the choices originally restricted to 'dramas' and 'novels'. With Romanticism, 'poetry' came to be included as well. Hence, at the end of the nineteenth century, under the influence of literary criticism and mass-produced literature, subgenres and mixtures of the three classic forms were accepted as genres and a comprehensive system of defined kinds of literary writing was established.

This normative understanding prevailed until the late 1960s, when it was challenged from within literary theory and by other disciplines, such as anthropology, pragmatics, and text theory. 'Genre' as an exclusive literary concept was rejected. Genres were simply text types and were approached descriptively. However, different forms of opera, film, video, music, dance, professional and everyday practice—in brief, *any kind of cultural communication*—were soon considered genres as well. Genre had returned to its origin, 'kind'.

Genre had traditionally often been defined one-sidedly by proceeding from one of the three main aspects of text: form, content, or use. For instance, formal (or syntactic) orientations would argue that genres such as limericks and death notices could be recognized from certain textual structures. Contentual (or semantic) views would focus on typicality in content, seeing patents and psalms, for instance, as particular kinds of descriptions. Finally, focusing on use (or function), genres were primarily seen as (social) action, such as commands and recipes.

As long as linguistics focused mainly on pronunciation, grammar, and meaning at the sentence level or below, higher levels, such as text and context, were of less interest. However, once linguistic pragmatics started focusing on the use of language, perceptions of genres went through a paradigmatic change. Genres were now seen as function-oriented tools for social action. This created a tension between a functionalist/pragmatic and a semantic/syntactic approach. In the 1980s, there was successful advocacy for a syntactic–semantic approach to genre within the field of film/media, with both form and content defining genre. In other words, the relationship between the main communicational aspects of genre was unresolved.

More recently, however, language and communication are increasingly viewed as multimodal and systemic. Hence, genres are often perceived as the interactive interplay of form, content, and use. Consequently, discourse and hence genres interrelate three major aspects of the lifeworld: 'uttering selves', 'referred worlds', and 'acting societies'. This triadic perspective considers genre as a recognizable balance of aesthetics (form), epistemology (content), and ethics (use). However, this multidimensional view raises new questions: Should genres be defined by their degree of stability? Can they be understood and analyzed one by one in isolation? Where are they really located? Can genres be defined properly?

To address the questions in order, genres are certainly not just given or static. Studying forms of communication historically and sociologically makes it clear that genres come into being, live, change, and fade out. Genres have to mix cultural stability and change. Hence, genres are now more often seen as relatively stable and flexible, some more stable than others, the variation motivated by social needs for openness and closeness. Further, it is clear that genres are inherited and create system-like connections. Narratives, from short jokes to 'endless' television series, for instance, share some aspects but differ in others. Genres are thus systemic, i.e. no aspect of a particular genre can be analyzed without defining its relationships with others.

Whether a verbal genre is *in* a text is disputed. The understanding of language as a describably closed system is often combined with the view that genres are repetitive, fixed macropatterns of communication located *in* texts. Approaches seeing genres as flexible and open phenomena, however, perceive them more as a set of potential forms, meanings, and uses. Metaphorically, genres can be seen as nine tenths of an iceberg, where the visible part (the utterance) only symptomatically hints or directs communicators to a potential intended meaning framed by a possible genre. Hence, even a verbal genre is a growing, immanent potential of mind and body, not just a kind of text; and, by the same token, a text is not *in* a verbal genre.

The relation between the two planes is regulated by a mechanism called, in text theory, theme/rheme, which balances the given (or the expected in relation to the genre, the context, or the situation) and the new: *Once upon a time there was* (theme/given in the genre 'fairy tale') *a big troll* (rheme/new). The relationship

between utterance/text and genre/context is dialogical in the sense that the genre influences the utterance. However, once an utterance is uttered, it contributes in principle to changing the genre. Over time, genres are changed by utterances.

The unpredictable and growing hidden and immanent potential makes it hard, if not impossible, to describe the 'true' nature of genres. Hence, a dilemma for applied linguistics is that focusing on stability may yield a more trustworthy and precise scientific discipline, but its value for application and practical understanding will be reduced. Some academic disciplines favor precise definitions to obtain valid operationalization in research. This has led either to a rejection of the concept of genre or to a predisposition for searching only for fixed aspects and hence to a self-fulfilling hypothesis of genre as fixed.

Thus, when the concept of genre has at the same time become more general in daily life as well as in the perception of several academic professions, a trustworthy and overall usable definition of the concept is endangered. However, while the acceptance of too much openness may lead to an increased understanding of the dynamics of text and context, it may reduce its scientific validity. Thus, genre urges analysts to decide whether the purpose of definitions and explanation is proof or understanding, precision or likelihood, description or application.

References

Altman, Rick. 1986. A semantic/syntactic approach to film genre. Film genre reader, ed. by Barry Keith Grant. Austin: University of Texas Press.

Altman, Rick. 1999. Film/Genre. London: British Film Institute Publishing.

Bakhtin, Mikhail. 1986. Speech genres and other late essays. Austin: University of Texas Press.

Ben-Amos, D. (ed.) 1976. Folklore genres. Austin: University of Texas Press.

Coe, Richard, et al. (eds.) 2001. The rhetoric and ideology of genre: strategies for stability and change. Hampton Press: Cresskill, New Jersey, USA.

Cope, Bill, and Mary Kalantzis (eds.) 1993. The powers of literacy: a genre approach to teaching writing. London: The Falmer Press.

Dubrow, Heather. 1982. Genre. London and New York: Methuen.

Fowler, Alastair. 1982. Kind of literature: an introduction to the theory of genres and modes. Cambridge, MS: Harvard University Press.

Freadman, Anne. 1994. Anyone for tennis? In Freedman and Medway.

Freedman, Aviva, and Peter Medway (eds.) 1994a. Learning and teaching genre. Portsmouth: Heinemann Boynton/Cook.

Freedman, Aviva, and Peter Medway (eds.) 1994b. Genre and the new rhetoric. Critical perspectives on literacy and education. London: Taylor and Francis.

Habermas, Jürgen. 1984. The theory of communicative action, Vol. 1, translated by T. McCarthy. Boston: Beacon.

Hauptmeier, Helmuth. 1987. Sketches of theories of genre. Poetics 16. 397–430.

Hirsch, Edward D. 1967. Validy in interpretation. New Haven: Yale University Press.

Martin, Jim. 1997. Analysing genre: functional parameters. Genre and institutions. Social processes in the workplace and school, ed. by Frances Christie and Jim Martin. London and Washington: Cassell.

Miller, Carolyn. 1994. Genre as social action. In Freedman and Medway.

Ongstad, Sigmund. 2001. Genres—from static, closed, extrinsic, verbal dyads to dynamic, open, intrinsic semiotic triads. In Coe et al.

Paltridge, Brian. 1997. Genre, frames and writing in research settings. Amsterdam: John Benjamins.

Swales, John. 1990. Genre analysis. English in academic and research settings. Cambridge, UK: Cambridge University Press.

Threadgold, T. 1989. Talking about genre ideologies and incompatible discourses. Cultural Studies 3(1). 101—27.

Todorov, Tzvetan. 1990/1978. Genres in discourse. Cambridge, UK: Cambridge University Press.

Voloshinov, V.N. 1973. Marxism and the philosophy of language. New York: Seminar Press.

SIGMUND ONGSTAD

Georgian and Caucasian Languages

The Caucasus is a range of mountains that run from the northwest near the Sea of Azov at the top of the Black Sea in southern Russia in a nearly straight line southwestward to the Caspian Sea coast where they end in the Apsheron Peninsula near Baku in Azerbaijan. Known as the Mountain of Tongues, the Caucasus Mountains harbor roughly 50 languages, at least 37 of which are indigenous to the region.

Georgian

Among the best known of the native languages is Georgian. Georgian belongs to the Kartvelian or South Caucasian language family, to which Mingrelian and Laz also belong, often grouped together as Zan or Chan, and the distantly related Svan in the mountains themselves. The family is located, as one of its names

suggests, on the southern slopes of the Caucasus massif, extending down toward the lowlands along the eastern Black Sea coast and well into eastern Turkey. Georgia is a republic and has roughly 3.5 million inhabitants who speak Georgian or a related language. The 1989 Turkish census listed 8.6 million people of 'Georgian' ancestry, most in the eastern regions. A few immigrant communities exist in North America and Europe, as well as in Russia.

Georgian itself is a literary language that dates back to the fifth century CE, and has an older (khutsuri or clerical) script and a later modern one (mkhedruli or knightly). Old Georgian exists as a bible and in other older church writing. The modern language is a vehicle for a whole culture and has an extensive literature, which may be dated to the 12th century.

Georgian is a language of bewildering complexity, not merely for reasons of its grammatical richness, but also because of a high degree of irregularity exhibited by that grammar. The sound system has many features typical of a Caucasian language generally, with the addition of two features that make it typically Kartvelian. First, the system shows a three way contrast in the stop consonant series and a two-way contrast in the fricative series, taking the dental stops and affricates as examples: stops /t, d, t'/, affricates /c, ʒ, c'/, fricatives /s, z/. It also has a series of ejective consonants. Georgian stands apart from other Caucasian languages, as do other members of its family, in lacking lateral fricatives and affricates, and in lacking pharyngeal fricatives. Georgian also exhibits the peculiar "harmonic clusters" of Kartvelian. These are two series coronal stops and affricates that are made with simultaneous velar or uvular closure. Thus Georgian has the coronals in (1).

Velarized and uvularized coronals
a. /tkeši/ pouring rain /txa/, /tqa/ goat
b. /dgoma/ to stand /dɣe/ day
c. /t'k'bili/ sweet /t'q'e/ forest
d. /ckera/ to look /cxeni/, /cqeni/ horse
e. /ʒgera/ beat, pulse /ʒɣola/ to lead
f. /c'k'riala/ clear, cheerful /c'q'ali/ water
g. /čkmeta/ to pinch /čxiri/, /čqiri/ rod, stick
h. /ǯgup'i/ group /ǯɣabna/ to scribble
i. /č'k'ua/ mind, intellect /č'q'ap'i/ slush

If the harmonic (rounded) clusters are treated as unitary complex segments, then the syllable onsets of Georgian become relatively straightforward: C(onsonant)V(owel), CCV, CSV, and SCV, where S stands for any sonorant (consonants with acoustic properties similar to vowels, e.g. r, l, y, w, m, n). One can have great strings of consonants, such as /wprckwnd-i/ = [fp'rc' °k' °ndi] 1-peel-past-1.past, 'I peeled it,' (coll.),

which is CCSCCSCVC. In fact Georgian is typologically highly unusual in allowing such strings, but also admitting strings of vowels, as in /aairebadi/ 'transient, volatile, volatility,' or /da-a-aprak'-d-a/ direction-active-warp-past-3, "It warped it," with each /a/ given a slight pulse of energy. Georgian has the canonical pure vowels /a, e, i, o, u/, with only Svan showing vowel length. Georgian has a light percussive intonational stress, which is highly unusual in that it is variable, falling sometimes on the penultimate and at other times on the antepenultimate syllable, for example /tbilísi/, /tbílisi/. Despite its consonantism Georgian is mellifluous.

Georgian has very complex word forms and word formation processes. Derivational processes changing the class of a word, e.g. from verb to noun, generally proceeds by means of suffixation:

a. /tavisupal/ 'free'   vs.   /tavisupl-eba/ 'freedom' (with syncope)

b. /bawšw/ 'child'   vs.   /bawšw-oba/ 'childhood.'

Circumfixes, i.e. combinations of prefixal and suffixal elements that are used simultaneously, are also widely used:

a. /or/ 'two'   vs.   /me-or-e/ 'second'
b. /xšir/ 'frequent'   vs.   /si-xšir-e/ 'frequency'
c. /kartwel/ 'Georgian'   vs.   /sa-kartwel-o/ 'Georgia'
d. /ɣrma/ 'deep'   vs.   /u-ɣrm-es-i/ 'very/the most old'

Georgian nouns and pronouns do not distinguish gender, but are marked for seven grammatical cases. These encode grammatical relations (ergative, absolutive, dative), address (vocative), possession (genitive), means (instrumental) and manner (adverbial).

Most Georgian morphology takes place in the complex verb, which can consist of up to ten morphemes, with many irregularities. Georgian verbs can express highly complex meanings for which other languages need whole sentences, as the following examples show – each given with a rough literal translation and a possible free translation:

a. ga-v- c'itl-d-e-t
   surface- I- blush- start- maybe- plural
   '…(that) we might blush'
b. ga-mo-gw-e-č'ri-a surface-
   toward.speaker-us-medial-wound-was
   'We wounded him.'
c. mo-m-k'lav-en
   toward speaker me kill they
   'They will kill me.'

d. s- daraʒob-s
   him watch he(subject)
   'He was watching him'

Georgian shows ergative – absolutive case markings for subject and direct object with verbs of completed action, absolutive – dative markings for verbs of incomplete or future action, and dative – absolutive for perfect tense verbs. Determining how this rich system of case marking interacts with more conventional aspects of syntax is a central problem in the study of Georgian.

Nakh-Daghestanian Languages

To the east of the Kartvelian family lies the Nakh-Daghestian or Northeast Caucasian family. This consists of roughly thirty-two languages. See the table, where * denotes literary languages, all written in modified Cyrillic and, with the exception of scant material, all dating from the Soviet period. They extend from the center of the northern Caucasus down to Azerbaijan, where they trail off into isolated villages. Hinukh may be the smallest speech community on earth.

Nakh-Daghestanian languages (approximate population)

1. Vai Nakh *Chechen (792,000), *Ingush (197,000), Batsbi (Kisti) (3,000)
2. Avaro-Andi-Tsez
   a. *Avar (7 dialects) (501,000)
   b. Andic – Andi (9,000), Karata (5,000), Chamalal (4,000), Bagwalal (4,000), Akhwakh (5,000), Botlikh (3,000), Godoberi (2,500), Tindi (5,000), Udi (formerly 6,000),
   c. Tsezic - Tsez (Dido) (7,000), Hinukh (200), Hunzib (400), Khwarshi (1,000), Bezhta (Kapuchis) (3,000), Inkhokwari (unknown)
   d. Lakk-Dargwa - *Lakk (92,000), *Dargwa (282,000), Kubachi (3,000), Khaidaq (28,000),
   e. Lezghian - *Lezgi (367,000), *Aghul (14,000), *Tabasaran (78,000), *Rutul (15,000). Archi (1,000), *Tsakhur (19,000), Budukh (1,000), Khinalug (2,000)
3. Kryz (6,000)

These languages contrast plain stops with "intense" ones (marked with full colon). The numerous lateral affricates are velar in their articulation. Contrasts occur among velar, uvular, pharyngeal, and laryngeal consonants. Some, such as Archi and Aghul, add pharyngealized uvulars as well. In many languages the presence of a pharyngeal or pharyngealized sound element in a word triggers the spread of pharyngealization across the word. The Richa dialect of Aghul is the only language known to contrast pharyngealized uvulars, true pharyngeals, "adytals" (sounds made with the epiglottis covering the adytus, the opening of the larynx), and laryngeals.

Richa Aghul contrasts
a. uvular /xǎw/ 'house' /ɣad/ 'hammer'
b. pharyngealized uvular /x̄ǎw/ 'nut' /ɣ̄ǎb/ 'stack'
c. pharyngeal /ḥ̣aw/ 'udder' /ʕ̣an/ 'belly'
d. adytal /ḥač/ 'apple' /ʕak°/ 'light'

The vowel systems are also rich. While the Lezghian and Lakk-Dargwa languages lack /o/, the others can have any number of vocalic contrasts, including front rounded and high back unrounded vowels. The vowels can also come nasalized, long, and adytalized, the last having a harsh quality. Dargwa shows an opposition between truly pharyngealized vowels and adytalized ones, the only language on earth to show this in its vowels. Godoberi (an Andic language) has "stiff" vowels, made with great articulatory force. Vocalic sequences are prohibited.

The morphology of these languages is characterized by a grammatical class system, ranging from the familiar gender system of male, female, neuter, and plural, to systems (in Vai Nakh) with six classifications, both in singular and in plural, or with fifteen in Godoberi, with masculine, feminine, and neuter singulars, and with human and neuter plurals, each with three distinct sub-types. The Lezghian languages tend to lack any grammatical class. A few nouns will exhibit these markers, for example Avar /w-ac:/ masculine(masc)-sibling, 'brother', /y-ac:/ feminine(fem)-sibling, 'sister', /b-ac:/ animal-sibling, 'littermate of an animal'. Generally, however, such markers occur as prefixes or suffixes on pronouns, adjectives, associated adverbs, and verbs, which reflect the class of the noun in the absolutive, as for example with Avar /hán-iw do-w ha-w-ize/ here-adverb masc 3rd-masc be.born-masc-infinitve, '(for) him to be born here'.

The languages tend to have a rigidly S(ubject)-O(bject)-V(erb) word order, with adjectives and relative clauses preceding the word they modify. Only in Vai Nakh do OVS and VSO orders occur in certain discourse environments. Embedded clauses generally precede the main clause. Some examples from Avar:

a. /w-as w-ekér-ula/
   masc-child masc-run-he.now
   'The boy runs.'
b. /čí t'axá-l r-ós-ul-e-w w-úgo/
   man book-plural plural-buy-now-masc masc-is
   'The man is buying the books.'
c. /ḥinč'.-a-ƛ·u-l mac'· b-úgo du-r, k'aƛa-l-é-ƛ·u-l/
   bird-its-animal voice animal-is you-of,

speak-now-of 'When you speak you sound like a bird.' Literally, 'Bird's voice is yours, while (you are) speaking.'

Northwest Caucasian Languages

The Northwest Caucasian family lies to the west of Kartvelian, where Abkhaz adjoins Mingrelian, and extends up along the Black Sea coast to the Sea of Azov, thence eastward toward the center of the North Caucasus. Perhaps half a million speakers still live in this homeland, but ten times that number reside in a diaspora that has communities in Turkey, the Middle East, Europe, and North America. The languages are Circassian, in western (Adyghean) and eastern (Kabardian) forms, with the moribund Ubykh transitional to Abkhaz and Abaza (on the northern side of the massif from Abkhaz). Apart from some tales recorded in the 19th century, and perhaps some ancient monuments in cuneiform or in runes, the literary languages come from the Soviet period. All are written in modified Cyrillic except for Ubykh texts, which are written in Latin script. These languages are famed for the complexity of their consonantal systems and the simplicity of their vocalic one. Ubykh, for example, has 81 consonants and only two vowels. These languages can show aspirated and ejective fricatives, as well as pharyngealized labials and uvulars, palatalized uvulars, and labialized dentals, [pt], etc. Consontatal clusters abound, but vocalic sequences are absent.

Derivational morphology proceeds by suffixation and compounding. What is striking is the limited number of basic roots, only several hundred. Most of the vocabulary is made of compounds, often for terms that would seem basic. Some forms are extremely elaborate, as with Bzhedukh West Circassian /śha-a-nə-ɣˠə-pč a-y-a-šˁʸ-a-a-ƛ'-a/ head-/a/-purpose-door-direction-to-lead-in-/a/-near-to, 'shutter for a window.' In some cases the semantic machinery at work is clearly visible.

Progressively abstract compounds (Bzhedukh West Circassian)
a. /pqə/ 'bone, frame, strut'
b. /śḥa-pq/ head-bone, 'skull'
c. /ʔa-pq+ƛa-pq/ arm-frame+leg-frame, 'body' (as a structure)
d. /wəna-pq/ house-frame, 'framework of a house'
e. /waradə-pq/ song-frame, 'melody'

It is the verb in these languages that is one of the wonders of morphology. It can bear morphemes expressing (roughly from start to end) the absolutive noun, the speaker's interest in the action or incredulity, as well as the syntactic function of the complex (subordinated, adverbial, gerundive), any deictic setting for the action, benefactives and adversatives, other adjuncts, all the arguments of a sentence and their number, capability or inability, the geometry of the scene of action, details of motion to or from or around that scene, any incorporated nouns that modify the sense of the verb, causative agent(s), valence of the root, reiteration of a state or action, adverbs, tense, mood, number of the absolutive noun, adverbial and modal suffixes, and subordinating complementizers at the very end. It is more a phrase than a simple lexical unit. Discourse often consists of strings of verbs with only the occasional noun:

Bzhedukh West Circassian verbal morphology
a. /sa sə-qə-z-a-tˤ-y-a-y-ə-ɣa-wəćˁə-žʸə-žʸə-ɣa/
I me-change.of.state-self-dative-surface-direction-to-3-past-let-stop-again-finally-past
'He let me stop again at last.'
b. /ø-qə-z-fə-w-śˀˁə-z-a-tⁿ-y-a-wəćˁə-śˀə-aḥe-ɣa-gˠara-r/
3-change.of.state-what-for-you-despite-self-dative-surface-direction-to-stop-able-around-past-definite.gerund-absolutive
'why he was able to stop all around despite you/your efforts'
c. /sa a-r ø-qə-s-ʔa+čˀʸə-w-fa-s-wəčˀʸə-na/
I 3-absolutive 3-interest-my-arm+instrumental-you-for-I-kill-immediate.future
'I shall kill him with my own hands for your sake.'

Syntax is ergative with SOV order. Modifiers follow and embeddings precede heads.

References

Aronson, Howard I. Georgian, a Reading Grammar. Columbus, Ohio: Slavica, 1982.
Colarusso, John. A grammar of the Kabardian Language. Calgary, Alberta: The University of Calgary Press, 1992.
Ebeling, Carl L. "The Grammar of Literary Avar (A Review Article)," Studia Caucasica, 2: 58–100, 1966.
Haspelmath, Martin. A Grammar of Lezgian. Berlin, New York: Mouton de Gruyter, 1993.
Kibrik, Alexandr E. (editor). Godoberi. Munich: Lincom-Europa, 1996.
Kibrik, Alexandr E. and Sergo V. Kodzasov. Sopostavitel'noe izuchenie dagestanskikh jazykov, imja. fonetika. Moscow: Moscow University Press, 1990.
Schultze, Wolfgang. Tsakhur. Munich: Lincom-Europa, 1997.
Tschenkeli, Kita. Einführung in die Georgische Sprache, band 1, theoretischer Teil. Zurich: Amirani Verlag, 1958.
Tuite, Kevin. Svan. Munich: Lincom-Europa, 1997.
Van den Berg, Helma. A Grammar of Hunzib. Munich: Lincom-Europa, 1995.

JOHN COLARUSSO

German

German is one of the main cultural languages of the Western world, spoken by approximately 100 million people around the world. It is the supraregional standard of a connected language area in Central Europe and the national language of both Germany (77 million) and Austria (8 million), and it is one of the four official languages of Switzerland, spoken by three fourths of the population (4.5 million speakers). Additionally, it is spoken in Alsace-Lorraine in eastern France, in the region of Alto Adige in Italy, and also in eastern Belgium, Luxembourg, and Liechtenstein. German forms a continuum from Switzerland north to the sea; a local dialect can be understood by speakers of nearby dialects, but not necessarily by speakers of far-away dialects.

Reliable statistics are not available concerning the number of German-speaking persons who inhabit those regions of Eastern Europe from which the Germans were expelled at the end of World War II: regions of the former Soviet Union (especially Ukraine), Hungary, the Czech Republic, Slovakia, Poland, and Romania (all of them former regions of the Austro-Hungarian Empire). Outside Europe, the largest number of people using German as their mother tongue live in the United States (1.5 million). An important group of German-speaking people in the United States speak the so-called Pennsylvania Dutch. German is also spoken in Australia, Canada, Argentina, Brazil, Chile, and in former German colonies (Namibia, Congo, Cameroon). German is important as a cultural and commercial second language for millions of people in Central, Northern, and Eastern Europe and in North and South America.

History of the German Language

German is a member of the Western Germanic family. A descendant of the Old High German language, German is a mixture of dialects that, following the so-called 'High German Consonant Shift' (sixth to eleventh centuries) that affected the southern areas, were divided into Low German (*Niederdeutsch, Plattdeutsch*), spoken in the lowlands of the North of Germany, and High German (*Hochdeutsch*), spoken in the highlands of the South of Germany and Austria. The term 'Low German' is essentially a geographic one, referring to the coastal, or lowland, area of the German region, as opposed to the High German area. The usually cited dividing line, south of which High German is spoken, runs eastward from Aachen, south of Düsseldorf, Kassel, Magdeburg, and Berlin, to Frankfurt-an-der-Oder. Although dialectal differences within both the High German and Low German regions remain, a trend toward uniformity in the direction of the written standard is expected partly as a result of widespread broadcasting, diminishing isolation, and increased socioeconomic mobility.

High German is the standard written language, even in the regions where Low German is more commonly spoken; High German is also the standard used in the schools of Austria and Germany, taught to foreigners, and used in the worlds of business and communications. High German is divided into Upper German, in Switzerland, Austria, Liechtenstein, and southern Germany, and Middle German, across Luxembourg and the middle of Germany.

Historically speaking, German falls into three main periods: Old German (750 CE–1050 CE), Middle German (1050–1500), and Modern German (1500 up to the present). The earliest existing records in German date back to about 750 CE. At that time, the language was developing in the Kingdom of Franks, especially in its eastern parts where the population was formed mainly of Germanic tribes. Dialects of this region formed a single language around the eighth century. Old German is characterized by the use of local dialects in writing and by the absence of a standard language. During the Middle German period, a relatively uniform written language developed in the government: during the reign of Louis IV (1314–1347), the Holy Roman emperor, German was adopted as the language of official court documents, instead of the Latin that until then was dominating official writings. Middle German had additional syllable vowels and a more analytic syntax. Besides, it was the first common language of the people that was the language of chivalric poetry. Between 1480 and 1500, it was introduced for official use in many municipalities and in the courts of Saxony and Meissen and was also adopted by the universities of Leipzig and Wittenberg. By 1500, German had generally been accepted as the official language in all parts of Saxony and Thuringia. It was this German that Martin Luther, often referred to as 'the creator of New High German', adapted for his translation of the Bible in 1523. His translation had such a wide distribution and acceptance that the East Middle German dialect in which it had been written came to serve as the authoritative exemplification of

modern High German. In addition, the publication of books in German developed in the East Middle German towns of Wittenberg, Erfurt, and Leipzig, as well as in such western and southwestern cities as Mainz, Strassburg, Basel, Nuremberg, and Augsburg. These developments helped reduce regional differences and standardize the literary language.

Standard written German emerged during the first quarter of the sixteenth century in the eastern midland area of Erfurt, Meissen, Dresden, and Leipzig, where the inhabitants, originally from regions farther west and southwest, spoke a dialect based on the Middle and Upper German dialects of High German. Despite the progress made in the sixteenth and seventeenth centuries to develop and maintain a *Nationalsprache* (national language), French still continued to be a considerable influence on the German language because it was the language of the German courts. During the reign of France's 'sun king', Louis XIV (1643–1715), it seemed that French was very much in fashion with the upper and middle classes. Not surprisingly, nevertheless, several societies began to develop a tendency against the influence of foreign languages on German, more particularly dedicated to the preservation and purification of the German language.

By the eighteenth century, German-speaking lands had generally adopted a written language, even though common spoken language continued to be marked by local and regional variations.

As for Low German, its history started in the thirteenth century as the heir of Old Saxon (Old Low German). Formerly a flourishing language, spoken in the rich cities of Hansa, it later suffered the strengthening influence of High German. Nowadays, it is used mostly in rural areas, although some revival has been taking place for the last 50 years in the universities of Northern Germany. The Low German vocabulary has generally remained much more archaic. In the last two centuries, there have been considerable loanwords of High German origin, especially in the field of culture and industry.

The history of the German language was affected by several systematic shifts of certain consonants. The so-called Germanic consonant shift distinguished the ancient Proto-Germanic language from other Indo-European members. In this shift, described by Grimm's Law, Indo-European *p, t, k* changed to Germanic *f, th, h,* respectively; Indo-European *b, d, g* to Germanic *p, t, k*; and similarly Indo-European *bh, dh, gh* to Germanic *b, d, g.* After Western Germanic had developed its own distinctive traits, the High German sound shift occurred (500–700 CE): the Germanic *p* became *pf* under certain conditions (High German *Pflanze,* Low German *Plante,* 'plant'); when used medially or finally after vowels it became *ff* or *f*

(High German *hoffen,* Low German *hopen,* 'to hope'). Another consequence is the transformation of the Germanic *t* in *z* (*Pflanze*) or *ss* (High German *essen,* Low German *eten,* 'to eat'). After vowels, *k* became *ch* (High German *machen,* Low German *maken,* 'to make'); in all other cases, *k* remained unchanged except in the extreme south of Germany, where it first became *kch,* and later *ch.* A later change, also found in Low German, is that of the Germanic *th* to *d* (High German *das,* Low German *dat,* 'that').

Characteristics of German

Another characteristic of German, as well as of all the Germanic languages, is that the principal accent falls regularly upon the first syllable of a word; in verbal combinations, however, the root syllable, not the prefix, is stressed (see examples below).

The German language makes extensive use of inflectional endings, both for nouns and verbs. However, Old High German was much more inflected than Modern German. Before the tenth century, its speakers used numerous endings: the noun could vary in gender (three genders), case (five cases), and number. Over time, this complex system of inflections was simplified.

Modern German nouns are inflected for case and number, but not for gender. Of 43 nominal inflections that existed in Old High German, only nine survived in the modern language. Still, the German language remains much more inflective than its relatives English or Dutch.

German nouns are classified according to three genders (masculine, feminine, and neuter). The gender is most easily identified by the noun's definite article in the nominative case: *der* (masculine), *die* (feminine and plural), and *das* (neuter). Sometimes, the grammatical gender of a noun has little to do with biological gender: *die Sonne* 'the sun' is feminine, but *der Mond* 'the moon' is masculine. There is a gender distinction for German nouns in the singular, but there is no gender distinction in the plural. All plural nouns (in the nominative and accusative) take the definite article *die*.

Nouns can appear in four cases: nominative (indicating the subject of the sentence), accusative (the object), dative (the indirect object), and genitive (the possessor of any other noun). The last three cases are also used after prepositions, according to very specific German grammar rules. This case system allows for a freer word order, so that subject, object, and indirect object can be placed in any order in the sentence. This flexibility of word order also allows for word play and greater poetic license.

The function of a noun in a German sentence determines its case, which, together with its gender, determines the article, and the endings on modifiers and

adjectives preceding the noun (or on the noun itself): e.g. *Der junge Mann sitzt hier.* 'The young man is sitting here.' (nominative); *Kennen Sie den jungen Mann?* 'Do you know the young man?' (accusative); *Ich gebe es dem jungen Mann.* 'I give it to the young man.' (dative); *Wie ist der Name des jungen Mannes?* 'What is the young man's name?' (genitive).

There are seven types of pronouns in German: personal pronouns (e.g. *ich* 'I', *wir* 'we', etc.), reflexive pronouns (e.g. *Sie wäscht <u>sich</u>.* 'She washes herself'), demonstrative pronouns (with the same form as the definite articles), indefinite pronouns (e.g. *viele* 'many', *etwas* 'something', etc.), relative pronouns and interrogative pronouns (e.g. *wer* 'who', *wen* 'whom', *wessen* 'whose', etc.), and intensive pronouns.

The proper German personal pronoun for the second person depends upon the closeness to the person one is addressing: i.e. *Sie* and its related forms are used when addressing one or several persons whom you would normally call by the last name, while *du* and its related forms are used when addressing one person (*ihr* more than one person) whom you would normally call by the first name: e.g. *Herr Schmidt, kennen Sie meine Frau?* 'Mr. Schmidt, do you know my wife?', *Robert, hast du meinen Hut?* 'Robert, do you have my hat?', *Kinder, ihr seid zu laut!* 'Children, you are too loud!' Depending upon its context, the personal pronoun *sie* can have various meanings. When it is capitalized and does not begin a sentence, it always means 'you' (singular or plural in a formal way): e.g. *Was meinen Sie dazu?* 'What do you think of that? When it is not capitalized and appears as the subject of a singular verb, it means 'she' or 'it' if it refers to an object of feminine grammatical gender: e.g. *Was meint sie dazu?* 'What does she think of that?' When it is not capitalized and appears as the subject of a plural verb, it means 'they'. When it is used as the direct object of a verb, it can mean 'her' or 'it' (if it refers to an object of feminine grammatical gender), or 'them': e.g. *Ich fahre sie durch die Stadt.* 'I drive her (or them) through the town.'

There were several changes during the history of the German verbal system. During the Old High German period, the verb could take many different forms. Nowadays, the number of main forms of the verb is reduced to three: *helfen - half - geholfen* 'help - helped - helped' instead of Middle High German *helfen -half - hulfen -geholfen.*

There are four types of verb in German: *Hilfsverben* (auxiliary verbs), *Modalverben* (modal auxiliary verbs), regelmäßige Verben (regular verbs), and unregelmäßige Verben (irregular verbs). There are 173 irregular verbs in Modern German, but this class is decreasing in number. However, it still contains many verbs belonging to the basic vocabulary.

The only tenses the German verb can form by itself are *Präsens* (present tense), *Präterium* (past tense), *Konjunktiv* I (subjunctive I), and *Konjunktiv* II (subjunctive II). In all other tenses, one of the verbs *haben* (to have), *sein* (to be), *werden* (to become, to get) is used as an auxiliary verb, in conjunction with the *Partizip* II (past participle) or the *Infinitiv* (infinitive) form of the principal verb. Generally, German uses the verb *werden* 'to become' to form the future tense, the verbs *haben* 'to have', and *sein* 'to be' to form the perfect tenses. The verb *haben* is generally used with transitive verbs and *sein* with certain intransitive verbs. Nearly all of the verbs conjugated with *sein* are strong verbs or irregular.

There are six modal auxiliary verbs in German: *dürfen* 'to be permitted, may', *müssen* 'to have to, must', *können* 'to be able, can', *mögen* 'to like to', *wollen* 'to want to', and *sollen* 'to suppose to, ought to'.

German has a large number of phrasal verbs made up of a verb (e.g. *kommen* 'come') and one or two particles or prefixes (e.g. *an* 'at'). These particles and their verbs have the force of single-word verbs (e.g. *ankommen* 'arrive'). Many of these particles attached to verbs are also ordinary prepositions: *ab* 'away from', *ein* 'into', *vor* 'before', etc. They provide additional (or sometimes changed) meaning to the verb. Occasionally, some of these prefixes become separated from the verb in clauses. In a simple sentence, these prefixes are separated from the verb at the end of the clause: *Sie kommen nun an.* 'They are arriving now.' They join the verb in the past tense: *Wir sind angekommen.* 'We have arrived.' When an infinitive is used or if the verb appears at the end of a subordinate clause, the prefix remains attached: *Wir werden ankommen.* 'We will arrive.'

Certain verbs have inseparable prefixes that are not used alone, have no fixed meaning alone, and are never separated from the verbs they are accompanying. An example would be *be* in *bekommen* 'get.' Separable prefixes have intonational stress, as in <u>an</u>kommen, but prefixes are never stressed, as in be<u>kommen</u>.

Adjectives and adverbs have the same form in German. The context in which they are used indicates how to interpret them: *Dieser Käsekuchen ist sehr gut.* 'This cheesecake is very good' vs. *Sie bäckt den Kuchen sehr gut.* 'She bakes the cake very well.'

The major word classes are interconnected by a complex network of word-formation rules. A distinctive feature of German is its extensive use of lengthy compound words. For example, the German counterpart of the English 'history of antiquity' is *Altertumswissenschaft.*

Traditionally, German was written in a Gothic or Black Letter style of the Roman alphabet, known in

German as *Fraktur* (meaning 'fractured'), which dates from the twelfth century. In the period following World War II, however, *Fraktur* was largely superseded by the Roman characters exclusively used throughout the rest of Western Europe. The Roman script contains the symbol ß or double s, which is used only in the lower case and represents a voiceless *s* (as in English *mouse*), and the umlauted vowels *ä, ö,* and *ü*. German is the only language in which all nouns are capitalized, common as well as proper (as they were in English several hundred years ago). This makes the nouns easy to spot when reading a sentence. Verbs and adjectives functioning as nouns are also capitalized.

Pennsylvania German

Pennsylvania German, also known as 'Pennsylvania Dutch' or simply 'Dutch', is spoken by 300,000 native speakers, descendants of German colonists, who left the Palatinate region of Germany during the late seventeenth and eighteenth centuries and settled in the southeastern part of Pennsylvania in the United States. They speak the Rhine-Franconian dialect with relatively few admixtures of English. It is the vehicle of a typical folk literature and its speakers are all bilingual, as the official language is English.

References

Beedham, Christopher. 1995. German linguistics: an introduction. Munchen: Iudicium.

Engel, U. 1988. Deutsche Grammatik. Heidelberg, Julius Groos.

Hammond, Robin. 1981. A German reference grammar. Oxford: Oxford University Press.

Keller, R.E. 1978. The German language. London: Faber.

Lippi-Green, Rosina L., and Joseph C. Salmons (eds.) 1996. Germanic linguistics. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Russ, C.V.J. (ed.) The dialects of modern German. A linguistic survey. London: Routledge.

Russ, Charles C.V. 1994. The German language today: a linguistic introduction. London, New York : Routledge.

Stevenson, Patrick (ed.) 1995. The German language and the real world: sociolinguistic, cultural, and pragmatic perspectives on contemporary German. Oxford: Clarendon Press.

Waterman, J.T. 1966. The history of the German language. Prospect Heights; IL: Waveland Press, Inc.

LAURA DANILIUC

See also **Germany**

Germany

Germany, or more accurately the Federal Republic of Germany (*Bundesrepublik Deutschland*), consists of 16 states (*Bundesländer*), five of which constituted the German Democratic Republic (East Germany) until reunification in 1990. Germany is situated in Central Europe, with the Baltic Sea, the North Sea, and Denmark to the north, Poland and the Czech Republic to the east, Austria and Switzerland to the south, and France, Luxembourg, Belgium, and the Netherlands to the west. It has an area of 357,020 square kilometers and a total population of 82,536,700 (as of December 2002).

The official language of the country is German Standard German (as opposed to e.g. Austrian Standard German). The development of this supraregional written norm is a relatively recent phenomenon and a good example of language planning, i.e. the use of deliberate, often official, measures relating to the status and/or form of a language. At the end of the Middle High German period (c. 1050–c. 1350), the written standard was still Latin, and the orthography, grammar, and vocabulary of the German texts of that time were not standardized but rather were strongly influenced by substantially different regional spoken language forms. Such regional variation was a result of the High German Sound Shift (also termed the Second Sound Shift), which had taken place to differing degrees in different parts of the German-speaking community from the sixth to eighth centuries CE. In this sound shift, the voiceless stops /p/, /t/, /k/ changed to voiceless fricatives /f/, /s/, /x/ and affricates /pf/, /ts/, /kx/, and the voiced stops /b/, /d/, /g/ to voiceless stops /p/, /t/, /k/. The dialects in the south of Germany, termed Upper German (*Oberdeutsch*) in keeping with the topography of the area, were affected first and most completely. Gradually, the shift moved northward. It became weaker as it did so, however, and the dialects in the center of Germany (Central German, *Mitteldeutsch*) were only partially affected, while those in the north (Low German, *Niederdeutsch*) remained unaffected.

During the Early New High German (Early Modern) period (c. 1350–c. 1750), however, the functional motivation for a standard increased, because of the development of towns and cities, the increase in

secular education, the invention of the printing press, and, in particular, because of Martin Luther's translation of the Bible into a variety of German intended to be understood by all speakers of German (1522–1534). Aside from these influences, a conscious drive was made by grammarians and lexicographers in the sixteenth century toward standardization, so that this process is thought to have been largely complete by the mid-eighteenth century. The variety of German that ultimately served as the basis for the development of written Standard German (High German, *Hochdeutsch*) was based on East Central German. Without doubt, the adoption of the standard was gradual, so that Charles Ferguson's concept of 'diglossia', i.e. a situation in which two varieties have mutually exclusive functions within a particular speech community, was relevant in the period following the introduction of the standard: speakers at that time would have used their local dialect (the Low variety) in the private domain of home and friends and the standard variety (the High variety) in the public domain.

The strict regulation of German orthography dates back to the mid-nineteenth century. Prior to this, there existed no unified spelling. Indeed, until a decree in 1862 ruled otherwise, various spellings could even be taught in a single school. The first significant attempts to standardize German orthography took place after the establishment of the German Reich (1871) at the *Verhandlungen der zur Herstellung größerer Einigung in der deutschen Rechtschreibung berufenen Konferenz* (Conference on the Establishment of Greater Standardization of German Orthography) in Berlin in 1876. The proposals of this conference, however, were never enforced. The process nevertheless continued, and at the second national orthography conference, held in Berlin in 1901–1902, a set of orthographic rules was developed on a trial basis for use in Germany. These were also adopted by Austria and Switzerland. The seventh edition of Konrad Duden's orthographic dictionary (*Orthographisches Wörterbuch der deutschen Sprache*, Orthographic dictionary of the German language), produced in 1902, incorporated these rules, and in December 1955, the education and arts ministers of the states of the Federal Republic of Germany awarded the Duden legal status, declaring it the orthographic norm for German. The rules of the second national orthography conference served as the orthographic norm for German until August 1998, when a spelling reform was instituted. The central motive for the reform process, which aimed at standardization and simplification, was educational, since it was argued that the time devoted to teaching spelling to children could be better spent otherwise and also that the complex spelling rules discriminated against working-class children whose exposure to written

German is generally less than that of middle-class children.

Adherence to the 'stem principle' is one of the rules introduced with the recent reform. Unlike in the old spelling, all consonants are to be maintained in the spelling of compound words, even in cases where three identical consonants may come together. The word for 'ballet dancer', a compound of *Ballett* 'ballet' and *Tänzer* 'dancer', thus becomes *Balletttänzer* with three *t*s rather than the former form, *Ballettänzer,* with two *t*s. Also, the decision as to whether to spell words apart or together was regularized. Rather than writing *radfahren* 'to ride a bicycle' but *Auto fahren* 'to drive a car', for example, both are now written apart, as *Rad fahren* and *Auto fahren.*

The reform process, which began in the 1950s, was a lengthy and complex one, clearly exemplifying the difficulties of language planning. Not only was inter-state and international consensus concerning the orthographic norm difficult to reach, but public support was also weak. It was not until several attempts at reform had been rejected that a list of recommendations (drawn up by a group of representatives from Germany, Austria, and Switzerland as well as Belgium, France (Alsace), Hungary, Italy (South Tyrol), Liechtenstein, and Romania) was accepted in July 1996 at an orthography conference in Vienna. These new spelling rules have represented the official norm since August 1, 1998. Their implementation and review is the task of the Institut für deutsche Sprache in Mannheim. The initial grace period during which both the old and new spelling may be used ends on July 31, 2005.

In public opinion, the spelling reform met with general rejection. Indeed, the state of Schleswig-Holstein, in a referendum on September 27, 1998, voted not to accept the reform. The reasons for the lack of public support were varied, although many involved emotion, laziness, or ignorance. Many individuals were reluctant to change since they identified with the old spelling, were not prepared to learn the new rules, or mistakenly equated language with spelling, believing that such planning would disturb the natural progression of a natural language. Other reasons involved the belief that the costs of the reform were not justified since the reform did not go far enough, representing a mere tidying-up process.

As for the standardization of German pronunciation, it was not until 1898 that the norm, codified by Theodor Siebs and based on the pronunciation of German Standard German in the north of Germany, was established (Stage German, *Bühnendeutsch*). However, although radio announcers from different parts of Germany may sound the same even today, regional pronunciations of the standard variety based

on the accents of the local nonstandard German dialects reign, since *Bühnendeutsch* was never completely adopted. Indeed, it is not only in pronunciation that German Standard German presently shows regional variation; German Standard German is colored with regional variation on every level of language, except the strictly regulated orthographic level. On the level of vocabulary, for example, the word *Samstag* is commonly used in the south and west of the country for 'Saturday', whereas *Sonnabend* is generally used in the north and east. Indeed, in recent years, the trend toward regionalism has become increasingly evident.

Despite standardization, the dialects of German are many and reflect the uneven effect of the aforementioned Second Sound Shift in various parts of Germany. The 'Benrath line' is an isogloss (i.e. a line drawn on a linguistic map to locate the outer limits of some characteristic feature) that runs from Aachen in the west through Benrath, a suburb of Düsseldorf, to Frankfurt an der Oder in the east, dividing the country into two regions. To the north are found Low German dialects not affected by the sound shift, and to the south, High German dialects affected by the shift. Dialects from the very north and very south of this divide are virtually mutually incomprehensible. In the transition zone in the middle (Central German), partial traces of the shift are found. Within the divide into Low, Central, and Upper German, further divisions are made based on particular linguistic criteria. The accompanying map provides a sketch of some of the principal differentiations. It must be pointed out, however, that none of these dialect areas are as homogeneous or as strictly defined as they appear (see Figure 1).

Standardization has led to an overall decline in the use of local dialects, and not all are equally healthy, despite a general trend toward regionalism. In the south, in Bavaria, Baden-Württemberg, and to a lesser degree in Hesse and parts of the Rhineland, dialects function as markers of regional identity to a substantial extent and are therefore relatively healthy, being used fairly far up the social ladder. In the north, however, their use is more socially stigmatized. Not surprisingly, therefore, German Standard German has already replaced many of the Lower German dialects of the north. Diglossia is rare in Germany, except in the case of elderly people living in rural parts of the Upper German dialect region and in remote rural areas in the Low German and Central German regions. Instead, most speakers today commonly shift their choice of variety along a continuum between the poles of dialect and standard, with most everyday spoken German being of a colloquial nature (*Umgangssprache*), a relatively new level of language that has developed in urban areas due primarily to the presence of a standard written variety. Colloquial German itself can be situated near or far from the standard, depending on the region and the specific situational constellations. In the north of the country, for example, colloquial speech is closer to the standard than in the south.

Germany is a multilingual country, and although German is the principal language, many other languages are also spoken in the Federal Republic. While many of these, such as Frisian, Sorbian, and Danish, are long-established minority languages, others, such as Arabic, Greek, Turkish, Hindi, Italian, and Spanish, were only introduced in the late 1950s via the immigration of guest workers (*Gastarbeiter*). In more recent years, there has also been immigration of individuals of German descent (*Aussiedler*) from Eastern countries. In 1999, 7.2% of the population were not of German nationality.

Frisian is the West Germanic language most closely related to English. Today, two mutually unintelligible main varieties are spoken in Germany: North Frisian and East Frisian. East Frisian, or more specifically Saterland Frisian, is still spoken in the Saterland, west of Oldenburg in East Frisia in the north of Germany (fewer than 1,000 speakers). It is being displaced by German. The same is also true of North Frisian (approx. 8,000–9,000 speakers), which itself includes nine very different, partially mutually incomprehensible, dialects spoken on the west coast of Schleswig-Holstein north of Husum and on the offshore islands of Föhr, Amrum, Sylt, Nordstrand, Pellworm, the ten islands of the Halligen group, and Heligoland. Reasons for the dwindling numbers of Frisian speakers are improved communication, the centralization of the education system, the large number of dialects of the language, and the fact that no standard written or spoken form exists. A form of colloquial language close to German Standard German is thus favored. Efforts to save the Frisian language from extinction have generally been rather insignificant. While the people of the tourist island of Sylt manage to nurture their identity and speak Frisian, speakers in other areas on the mainland have a more difficult task since they do not form an isolated group. In addition, many speakers wish to discard the 'socially backward and poor' image commonly associated with speakers of Frisian. Since the late 1980s, political engagement for the Frisian language has increased; money has been made available to teach Frisian on a voluntary basis in schools, for example. While such efforts may serve to delay the extinction of Frisian, it is doubtful whether they will prevent it.

Standard Danish is spoken by a recognized minority in southern Schleswig. Despite the existence of schools there in which Danish is spoken, the future for the language in this area is also rather pessimistic.
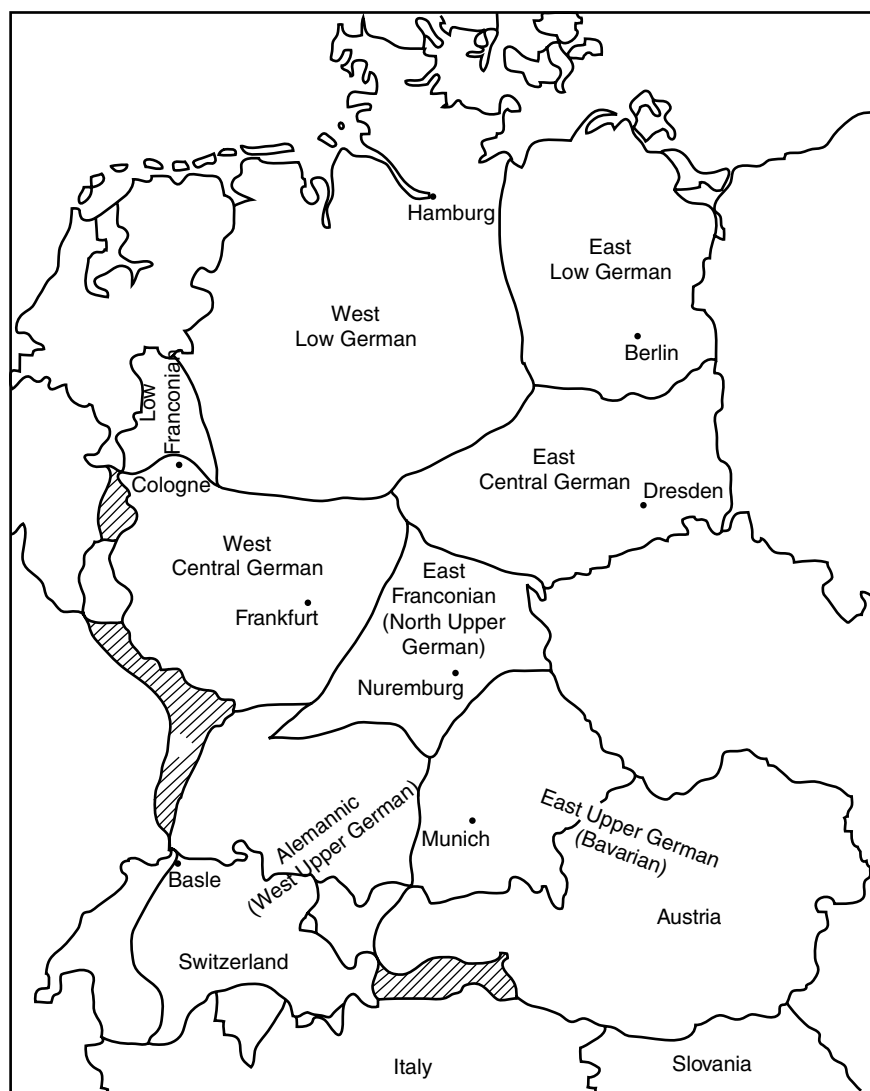
Figure 1. German Dialects.

Since the seventh century, the Slavic language Sorbian (approx. 67,000 speakers) has been spoken in Lusatia (Lausitz), a relatively vaguely defined area in former East Germany bordering on Poland and the Czech Republic and belonging administratively to the German states of Brandenburg and Saxony. Two varieties, Lower Sorbian and Upper Sorbian, are differentiated. Lower Sorbian is spoken in the area around Cottbus (*Chośebuz*) in Lower Lusatia and Upper Sorbian in the area around Bautzen (*Budyšin*) in Upper Lusatia. The Sorbians themselves never formed their own state but have remained under German rule. After a period of repression during the Nazi era, Sorbian was promoted in the German Democratic Republic. Its use in local government was authorized and children had the right to learn it in schools, although in reality it was often not possible to practice such rights. German reunification brought generous support for the language and culture of the Sorbian minority. In 1991, for

example, a foundation was established for the Sorbian people (*Stiftung für das sorbische Volk*) with the task of promoting the language and culture. However, despite such efforts, the survival of the Sorbian language, although the least endangered of the established minority languages in Germany, is uncertain since all speakers are bilingual in German, the tight bonds that existed between Sorbians in former times have dissolved in the meantime, and many young Sorbians living in German-speaking areas do not identify with their language.

As for foreign language teaching in German schools, not surprisingly, English, given its international status, is primary.

Gastarbeiter 'guest workers', also and more appropriately termed migrants, immigrants, work immigrants, and foreign workers, came to the Federal Republic primarily from countries such as Greece, Italy, Turkey, and the former Yugoslavia from the late

1950s onward, following a policy decision designed to ameliorate the acute labor shortage in West Germany at the time. They settled mostly in urban areas such as Munich, Stuttgart, and Frankfurt am Main, often in concentrated ethnic groups (e.g. Turks in West Berlin (Kreuzberg), Greeks in Munich, Yugoslavs in Stuttgart). The early 1970s witnessed an end to this policy, and immigration laws became stricter. With changing economic circumstances, these 'guest' workers were encouraged to return to their home countries from the early 1980s onward. But many have remained, settled down, and brought relatives to Germany. Unfortunately, this group leads a marginal existence in Germany, being denied dual citizenship, for example, on the basis of policy assumptions that this workforce was in Germany temporarily and that Germany is not an immigrant country.

Most first-generation foreign workers are not proficient in German despite having lived in the country for a number of decades. The majority of them had not enjoyed any instruction in German prior to their arrival and, since integration was not foreseen, few have ever been taught German systematically. With the exception of much of the Turkish population, though, many second-generation immigrants are bilingual, with German as the dominant language.

The linguistic study of the German spoken by the migrant workers has provided a number of insights into the process of uninstructed second-language acquisition. It has been demonstrated, for instance, that although speakers do transfer features from their native language to their second language, the German of speakers of languages as different as Spanish and Italian (Romance languages) and Croatian and Slovenian (Slavic languages) is characterized by similar simplifying measures, such as deletion and reduction. A typical example of the former is the utterance *Ich fahre zwei Wochen Irland* 'I'm going two weeks Ireland', in which two prepositions have been deleted—the utterance should read *Ich fahre **in/ für** zwei Wochen **nach** Irland* 'I'm going **to** Ireland **in/for** two weeks'. An example of reduction is the use of the informal '*du*' whatever the formality of the situation. Given such similarities between speakers of different native languages, it is possible to speak of the existence of Guest Worker German (*Gastarbeiterdeutsch*). Considerable variation is also present, though. An order of acquisition of grammatical structures has also been established based on the analysis of Guest Worker German. How far a particular individual progresses along the path of acquisition has been found to depend on a number of factors, the most important being the degree and intensity of contact with Germans, age on arrival, level of education acquired in home country, years of schooling, period of residence

in Germany, and the subjective social and psychological distance between the individual and native speakers of German.

Guest Worker German is also of interest to researchers of pidgins, i.e. languages without native speakers that arise out of a basic need to find a lingua franca (a common system of communication), since it shares many features of pidgins, even though it arises as a result of communication between migrant workers and German native speakers rather than between migrant workers of various cultures, the usual occasion for the development of a pidgin.

Finally, individuals of German descent (*Aussiedler*) from Eastern Europe, primarily from Romania, Poland, and the former Soviet Union, have been immigrating to Germany since the 1950s. However, such immigration, particularly from the former Soviet Union, increased dramatically between the fall of the Iron Curtain and 1995; since then, numbers have been steadily decreasing. While some of those from the former Soviet Union and Romania speak a variety of German, others from Russia speak an archaic dialect of German and can understand but not speak German Standard German. Those from Poland, on the other hand, have little or no competence in German. The German of these speakers shares many similarities with Guest Worker German.

In sum, the study of the linguistic situation in the Federal Republic of Germany is a rewarding endeavor. The study of the standardization of German Standard German sheds light on the difficult process of language planning, and the study of minority languages highlights the possible consequences of language contact, the issue of language shift, and, in the case of the immigrant languages, aspects of uninstructed second-language acquisition.

References

Ammon, Ulrich. 1995. Die deutsche Sprache in Deutschland, Österreich und der Schweiz. Das Problem der nationalen Varietäten. Berlin and New York: de Gruyter.

Barbour, Stephen, and Patrick Stevenson. 1990. Variation in German. A critical approach to German sociolinguistics. Cambridge: Cambridge University Press.

Clahsen, Harald, Jürgen M. Meisel, and Manfred Pienemann. 1983. Deutsch als Zweitsprache. Der Spracherwerb ausländischer Arbeiter. Tübingen: Narr.

Clyne, Michael G. 1995. The German language in a changing Europe. Cambridge and New York: Cambridge University Press.

Duden, Konrad. 1880. Vollständiges Orthographisches Wörterbuch der deutschen Sprache (= Duden), 1st edition. Leipzig: Bibliographisches Institut.

Duden, Konrad. 1902. Orthographisches Wörterbuch der deutschen Sprache (=Duden), 7th edition. Leipzig: Bibliographisches Institut.

Ferguson, Charles A. 1959. Diglossia. Word 15.

Heller, Klaus. Rechtschreibreform—Eine Zusammenfassung [HTML document], IDS (Institut für deutsche Sprache) Sprachreport Extraausgabe (July 1996), Retrieved [4.05.2001] from the World Wide Web: http://www.ids-mannheim.de/pub/sprachreport/reform/.

Langer, Nils. The Rechtschreibreform—a lesson in linguistic purism [HTML document], GFL: German as a foreign language [online] 3 (2000), Retrieved [4.05.2000] from the World Wide Web: http://www.gfl-journal.com/.

Russ, Charles V.J. (ed.) 1990. The dialects of modern German: a linguistic survey. London: Routledge.

Russ, Charles V.J. 1994. The German language today: a linguistic introduction. London and New York: Routledge.

Sauer, Wolfgang W., and Helmut Glück. 1995. Norms and reforms: fixing the form of the language. The German language and the real world: sociolinguistic, cultural, and pragmatic perspectives on contemporary German, ed. by Patrick Stevenson. Oxford and New York: Clarendon Press.

Stevenson, Patrick. 2002. Language and German disunity. A sociolinguistic history of East and West in Germany 1945–2000. Oxford: Oxford University Press.

ANNE BARRON

See also **Acquisition Theories; Diglossia; German; Language: Contact—Overview; Language Death; Language Planning; Standard Language**

Givón, Talmy

Talmy Givón's outstanding contribution to the study of language in a wide scientific setting cannot be understood without the influence of two leading figures of North-American linguistics: Dwight Bolinger and Joseph Greenberg. Like his teacher Bolinger, Givón has followed the central claim of the functional tradition of linguistics, which ultimately can be traced back to the Prague School of Linguistics and European functionalism, namely that language cannot be studied independently of its context. Like Greenberg, Givón has been concerned with describing cross-linguistic phenomena and finding functional explanations for them.

It is significant that among the many influences that Givón acknowledges in his works, which include fellow functionalists like Wallace Chafe, Bernard Comrie, Robert Dixon, John Haiman, Bernd Heine, Paul Hopper, and Sandra Thompson as well as Noam Chomsky, two teachers of Givón´s are mentioned with particular warmth and respect. The first is Bolinger, to whom Givón attributes an insistence that the form of language cannot be studied in isolation from its meaning, that grammar is a set of meaningful devices, and that the aim of the forms of language is to express thought. The other—rather unexpected—is a high school teacher of humanities who eventually became a reputed Egyptologist, Sarah Groll. Since Givón considers Groll´s teachings a formative experience that has lasted him a lifetime, it does not seem out of place to put down the interest in cultural and linguistic diversity shown by this author to the influence of Groll.

Along with these influences, three main principles could be said to characterize Givón´s work: the principle of antimodularity, the principle of cognitive plausibility, and the principle of linguistic relativity.

Considering antimodularity in the first place, it must be borne in mind that generative grammars, including formal approaches like Chomskian Transformational Grammar and functional approaches like Simon Dik and Kees Hengeveld´s Functional Grammar, consider grammar as a set of descriptive devices in such a way that several components or modules turn out different aspects of the final linguistic expression. The output of each module constitutes the input to the next module: typically, the lexicon turns out lexical items, the component of predication-formation fills in the semantic-syntactic slots of the clause with lexical items, thus turning out a sentence, and a morphophonological component gives the sentence its final (pronounced) form.

Cognitive plausibility implies that it is not enough to propose descriptive devices such as rules or operations, or even explanatory devices like specific or general principles. Much more than that, it is necessary to consider the plausibility of grammatical devices in the light of human abilities of memory and processing. For example, sentence elements or constituents bearing contrastive information, which typically occupy a clause-initial position in English, can be inserted into the correct slot by a rule that places contrastive focus constituents in the clause-initial position; or as a result of the application of a functional principle that stipulates that more important information is more likely to be placed earlier in the clause. For Givón, the cognitive basis of this functional principle is that the earlier a chunk of information is placed within its relevant unit, the more attention it attracts. Moreover, the information that attracts more attention is memorized, stored, and retrieved more efficiently. Another point of interest related to cognitive plausibility is Givón´s epistemology: for him, Western Philosophy cannot be reduced to

Plato and Aristotle. Kant, Peirce, and Wittgenstein also shed light on the social and communicative dimensions of language and, more importantly, on language categorization, which is not discrete but continuous, that is, based on the notion of prototype of a category.

Givón follows in Franz Boas and Edward Sapir´s footsteps in the question of linguistic relativity, advocating the study of each language on its own terms. The ideological basis of linguistic relativity is even more obvious in Givón than in the work of his predecessors: in order to avoid Eurocentrism (the scientific hegemony of Indo-European languages over languages belonging to other families), the conclusions reached after the study of English cannot be automatically generalized to all languages independently of their typological characteristics, and English cannot be the recurrent term of comparison for any linguistic topic. For this reason, Givón prefers a formally unconstrained grammar, which consists of functionally oriented principles, rather than a formally constrained grammar involving language-specific rules whose typological validity is, at least, questionable. Functionally oriented principles are based on explanatory devices like iconicity (or resemblance between the linguistic code and the extra-linguistic reality) and markedness (or asymmetry between two members of a pair or one of the members of a cluster and the rest of the members of such a cluster).

Givón has been a source of refreshing and thought-provoking inspiration; without him, some recent developments in functionalism would not be easy to explain.

Biography

Talmy Givón was born in Afula, Palestine on June 22, 1936. He received his B.S. in Agriculture in 1959; M.S. in Horticulture in 1962; M.A. in Linguistics in 1966; and Ph.D. in Linguistics in 1969. He taught at the UCLA Linguistics Department (1968–1981) and the University of Oregon Linguistics Department (1981–2002). He was Chair, Linguistics Department, University of Oregon (1981–1985) and Director, Cognitive Science Program, University of Oregon (1986–1987). He held visiting and consulting positions in Papua-New Guinea, New Zealand, Germany, Perú, Spain, and Japan. He was also Founding Editor *of Studies in African Linguistics* and *Typological Studies in Language* and Distinguished Professor (Emeritus) of Linguistics and Cognitive Science, University of Oregon.

References

Bates, Elizabeth, and Brian MacWhinney. 1982. Functionalist approaches to grammar. Language acquisition: the state of the art, ed. by Eric Wanner and Lila Gleitman. Cambridge: Cambridge University Press.

Butler, Christopher. 1991. Standards of adequacy in functional grammar. Journal of Linguistics 27.

Butler, Christopher, Ricardo Mairal, Javier Martín Arista, and Francisco Ruiz de Mendoza. 1999. Nuevas Perspectivas en Gramática Funcional. Barcelona: Ariel.

Givón, Talmy. 1979. On understanding grammar.

———. 1984. Syntax: a functional-typological introduction, Vol. I.

———. 1989. Mind, code and context: essays in pragmatics.

———. 1990. Syntax: a functional-typological introduction, Vol. II.

———. 1993. English grammar: a function-based introduction.

———. 1995. Functionalism and grammar.

Martín Arista, Javier. 1997. La adecuación psicológica de la gramática funcional tipológica. Cuadernos de Filología Inglesa 6(2).

Newmeyer, Frederick J. 2001. The Prague School and North American functionalist approaches to syntax. Journal of Linguistics 37.

Nichols, Johanna. 1984. Functional theories of grammar. Annual Review of Anthropology 13.

JAVIER MARTÍN ARISTA

See also **Functional Approaches; Generative Grammar; Greenberg, Joseph Harold**

Gothic

Gothic, the only documented member of the Eastern Germanic branch of the Indo-European family, is an extinct language once spoken by the Goths. The Gothic historian Jordanes (AD 551) says that his people came originally from Scandinavia; place names, including the island of Gotland, show their presence there around the first millennium BC. The Goths then moved from Scandinavia, through the Baltic islands, and to the Black Sea, where they divided into two major groups around AD 270.

The West Goths—the Visigoths—invaded the territory of the Roman Empire during the reign of Marcus Aurelius (AD 270–275) and settled in the province of Dacia (modern-day Romania) near the lower Danube

and then migrated to Gaul and Spain, where they established a powerful kingdom that lasted for over three to four centuries. Around AD 490 in Constantinople, the East Goths—the Ostrogoths—under the leadership of Theordoric established an Ostrogothic kingdom in Italy that was subsequently destroyed by the Byzantines in the middle of the sixth century. By the eighth century, both groups were assimilated into the larger populations who surrounded them.

Proto-Gothic, the direct predecessor of Gothic and one of the descendants of Proto-Germanic, can only be studied through probable reconstructions of Proto-Gothic words. However, early Gothic literary records that provide language samples for linguistic study are available and date from the fourth century, predating records of any of the other Germanic languages. Gothic is the cornerstone of Germanic linguistics because it provides the oldest extensive collection of written samples in the Germanic group of languages, except for a few brief Norse runic inscriptions. Comparable specimens of other Germanic language, like English, Frisian, High German, and Norse, do not appear until four to nine centuries later. Very little is known about the other Eastern Germanic languages of Burgundian, Gepidic, Rugian, and Vandalic.

Gothic had two main dialects: Ostrogothic was the language of the Ostrogoths in eastern Europe and Italy, and Visigothic was the language of the Visigoths of east central Europe, Gaul, and Spain. Some scholars believe that the Ostrogothic dialect became extinct after the fall of the Ostrogoth kingdom in Italy in the sixth century; however, Ogier Ghislain de Busbecq, a Flemish ambassador who served in Constantinople between 1560 and 1562, found traces of Gothic in the words and phrases of the area. Even though his scholarship and transcription were less than ideal, this variety is thought to be a dialect of Ostrogothic and is now known as *Krimgotisch* (Crimean Gothic). The Visigothic dialect is believed to have become extinct around the time of the Arab conquest of Spain in 711.

The earliest Gothic literary work and the cornerstone to all Germanic linguistics is the translation of the Bible by Wulfila (311–383), a Visigothic bishop from Moesia (modern Bulgaria). Wulfila's original translation of the Bible from Greek has not survived; however, several reproductions of his work from the fifth and sixth centuries still exist. The *Codex Argenteus*, written in the first half of the sixth century by Ostrogothic scribes in northern Italy, consists of translations of the gospels and epistles, providing about three-quarters of the New Testament. A bound volume of manuscripts, the *Codex Argenteus,* the most studied of the Wulfila reproductions, is written in gold and silver on purple velum. Discovered in the sixteenth century, it is now in the care of Uppsala University in Sweden.

Other Gothic texts are limited, but include the *Skeireins* (eight unconnected pages of a commentary on the Gospel of St. John), some deeds of sale, a fragment of a calendar of martyrs, the five *Codices Ambrosiani*, and bilingual manuscripts like the Gothic-Latin *Codex Carolinus*. However, the *Codex Argenteus* is the earliest literary remains in any Germanic language; the only earlier records are limited to a few Norse runic inscriptions and some loanwords preserved in a few non-Germanic languages.

Historians generally agree that Wulfila created the Gothic alphabet (which should not be confused with the Gothic script also known as Black Letter that is used by printers) with the sole purpose of translating the Bible. The Gothic alphabet has 27 letters; the majority of the letters were derived from Greek, but five or six of the letters came from Latin, and two were borrowed from runic script. In general, though, the Gothic alphabet resembles Latin and Greek scripts with some differences in the order and sounds of the letters. Gothic is written from left to right with only very rare spacing between letters. What we know about the pronunciation of Gothic is based on how the Greek and Latin foreign words and names in the Gothic Bible were transcribed.

Gothic, separated from other Germanic languages early on, does not exhibit many of the linguistic developments that appear in other Germanic languages in later periods. Therefore, the phenomena that are unique to Gothic are considered important to all Germanic linguistics.

One way in which Gothic can be distinguished from other members of the Germanic family is by its use of the long [ê] sound (as in *mênoþ* 'month'), a sound that is assumed to have been in Proto-Germanic and that disappeared in all other Germanic languages. Another Gothic singularity is how it preserved the original Proto-Germanic [z] sound (as in *huzd* 'hoard') that changed to [r] in all other Germanic languages.

Gothic nouns and pronouns had five Germanic cases: nominative, vocative, genitive, dative, and accusative; however, some of the pronouns and adjectives also retained the instrumental case that was lost in other Germanic languages. Gothic nouns, pronouns, and adjectives had three genders (masculine, feminine, and neuter) and three numbers (singular, dual, and plural). The dual number (a plural form expressing two persons or things) is an archaic feature that does not appear in later varieties of other Germanic languages. Gothic also had a reflexive pronoun (*seina, sis, sik* 'self') that can be traced back to Proto-Indo-European.

Some verb features specific to Gothic occur seldom, if at all, in the other Germanic languages that appear in later written samples. Germanic verbs, in general, can be divided into strong (regular) and weak (irregular) verbs, and each of these can also be divided into subclasses. The seventh class of Gothic strong verbs is distinguished from other Germanic languages in that it uses the process of reduplication (repetition of syllables) to form the verbs. For example, the past tense of the verb *háitan* 'to call' is formed by repeating the first consonant of the verb stem and attaching it after the vowel *ái*, resulting in a form that attaches as a prefix to the basic stem to form *haíháit* 'called'. Another verb feature unique to Gothic are the retained passive inflectional suffixes for verbs; all other Germanic languages use the periphrastic construction of auxiliary plus verb plus past participle (as in 'is called' in English). Gothic uses passive inflectional suffixes, as seen in the use of *-ada* to form the passive *háitada* 'is called' from the verb *háitan* 'to call'.

A sample Gothic sentence reads as *Þanuh þan in menoþ saihstin insandiþs was aggilus Gabriel fram guþa* ('then-then-in-month-sixth-sent-was-angel-Gabriel-from-God'), yielding the English translation 'Then in the sixth month was sent the angel Gabriel from God.'

References

Bennett, William H. 1972. An introduction to the Gothic language. New York: MLA.

Braune, Wilhelm, and Ernst A. Ebbinghaus. 1981. Gotische Grammatik, 18th edition. Tübingen: Niemeyer.

Krause, Wolfgang. 1968. Handbuch des Gotischen, 3rd edition. Munich: Beck.

Mierow, Charles C. 1915. The Gothic history of Jordanes. Princeton, NJ: Princeton University Press.

Penzl, Herbert. 1972. Methoden der germanischen Linguistik. Tübingen: Niemeyer.

Robinson, Orrin W. 2002. Old English and its closest relatives: a survey of the earliest Germanic languages. Stanford, CA: Stanford University Press.

ALLISON SMITH

See also **Indo-European 2: Germanic Languages**

Grammar, Traditional

Traditional grammar refers to the type of grammar study done prior to the beginnings of modern linguistics. Grammar, in this traditional sense, is the study of the structure and formation of words and sentences, usually without much reference to sound and meaning. In the more modern linguistic sense, grammar is the study of the entire interrelated system of structures— sounds, words, meanings, sentences—within a language.

Traditional grammar can be traced back over 2,000 years and includes grammars from the classical period of Greece, India, and Rome; the Middle Ages; the Renaissance; the eighteenth and nineteenth century; and more modern times. The grammars created in this tradition reflect the prescriptive view that one dialect or variety of a language is to be valued more highly than others and should be the norm for all speakers of the language. Traditional grammars include prescriptive rules that are to be followed and proscriptive rules of usage to be avoided. 'When describing an emotion, use of an English word descended from Latin is preferred over an Anglo-Saxon word' is an example of a prescriptive rule, and 'Never split an infinitive' is an example of a proscriptive rule.

The analytical study of language began around 500 BC in Greece and India. The work of Greek scholar Dionysius Thrax is the model for all grammars of European languages that follow. His *Hē grammátikē tékhnē* (c. 100 BC; The Art of Letters) was the first widely recognized text to provide a curriculum for learning proper Greek. His lessons included an introduction to the alphabet, lessons on how to join syllables together properly, and instruction in the appreciation of word arrangement (syntax). To Thrax, grammar was the technical knowledge necessary to produce the prestige language of poets, orators, and writers.

Around the same time, the Roman scholar Marcus Terentius Varro produced the 25 volumes of his *De lingua latina* (c. 100, About the Latin Language). Varro contrasted Latin with Greek, changed Greek grammatical terms into Latin, and formed his grammar of Latin by adapting Greek rules.

Other Latin grammars, influenced by the works of Thrax and Varro, were produced in the Middle Ages. Aelius Donatus published *Ars Grammatica* (c. fourth century, Art of Letters), and Donat Priscianus Caesariensis (Priscian) wrote *Institutiones grammaticae*

(c. sixth century, Grammatical Foundations), which is the only complete surviving Latin grammar.

As printing became more widely available in the Renaissance, European grammarians began the mass production of grammars of their languages by mirroring the Latin grammars of Varro, Donatus, and Priscian. These traditional grammarians presumed that the grammatical descriptions of Latin could be routinely applied to their own languages; this perception, however, was not accurate and resulted in many artificial prescriptive and proscriptive rules. Many of these false assumptions still carry over to attitudes about English today.

Continuing with this tradition, grammarians in the eighteenth century studied English, along with many other European languages, by using the prescriptive approach in traditional grammar; during this time alone, over 270 grammars of English were published. During most of the eighteenth and nineteenth centuries, grammar was viewed as the art or science of correct language in both speech and writing. By pointing out common mistakes in usage, these early grammarians created grammars and dictionaries to help settle usage arguments and to encourage the improvement of English.

One of the most influential grammars of the eighteenth century was Lindley Murray's *English grammar* (1794), which was updated in new editions for decades. Murray's rules were taught for many years throughout school systems in England and the United States and helped to create modern attitudes about the existence of a correct or standard variety of English. Murray's grammar represents a practice that continued to develop throughout the nineteenth century and was still dominant in the 1960s when linguistics began to focus more on generative and transformational grammar due to Noam Chomsky's groundbreaking and influential ideas.

Even though linguists today view traditional grammar as an unscientific way to study language and grammar, many of the basic Latin-based notions of grammar can still be found in all levels of the classroom and in textbooks and usage guides available to educators and the public. Traditional grammar books usually provide lists of grammatical terms, definitions of those terms, and advice on using so-called 'standard' grammar, including suggested correct usage of punctuation, spelling, and word choice. This advice is usually based on the prescriptive rules of prestige varieties of English, varieties often only able to be used by those in power either economically or politically.

Linguists, along with many English faculty, would rather have students study language with a descriptive approach that includes the analysis of real samples of a mixture of English dialect varieties, not just the prescribed, and sometimes inconsistent, prestige forms. Linguists or teachers using a descriptive approach say that it allows students to investigate language on a deeper level, enabling students to see the system at work, instead of teaching them isolated prescriptive and proscriptive rules based on Latin, a dead language no longer in flux as English constantly is.

Linguists also believe that the rules of traditional grammar are inadequate because many of the rules are oversimplified, inconsistent, or not consistently conformed to. The grammars of classical Greece and Rome were based on the best orators or poets of the day. However, the best poets or speakers of our day are lauded for their poetic use of language that breaks prescriptive rules. For example, a traditional grammar rule of modern English, often found in usage guides and student handbooks, forbids the use of fragment sentences like 'The train running up the hill.' However, e.e. cummings or Maya Angelou could use this sentence for poetic effect without question.

Many teachers themselves want to be trained in traditional grammar, even though its inconsistencies may not help them when they have to explain grammatical points to their students. The National Council for Accreditation of Teacher Education (NCATE) requires that teachers in training take linguistics or language courses to teach them to examine the differences between traditional grammar and more modern grammars. However, many English teachers view traditional grammar as necessary and newer grammars as little help to them. And even though more modern types of grammatical analysis exist, many students, future teachers, and the general public still believe grammar means the traditional Latin-based grammar of old.

References

Allen, Robert. 1972. English grammars and English grammar. New York: Scribner.

Battistella, Edwin. 1999. The persistence of traditional grammar. Language Alive in the classroom, ed. by Rebecca S. Wheeler. Westport, CT: Praeger.

Chomsky, Noam. 1965. Aspects of the theory of syntax. Cambridge, MA: MIT Press.

Haussamen, Brock. 1993. Revising the rules: traditional grammar and modern linguistics. Dubuque, IA: Kendall/Hunt.

Leitner, Gerhard (ed.) 1991. English traditional grammars: an international perspective. Philadelphia: John Benjamins.

Noguchi, Rei. 1991. Grammar and the teaching of writing. Urbana, IL: National Council of Teachers of English.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. A comprehensive grammar of the English language. London: Longman.

ALLISON SMITH

See also **Chomsky, Noam; Generative Grammar**

Grammar, Theories

Grammar theories have several overlapping main purposes in linguistics: to

- provide a model that clarifies our understanding of a particular language,
- model language in general,
- provide a convenient way of working with language, particularly with computers,
- describe the fundamental structures of language, and
- model human communication.

Noam Chomsky's transformational grammar has, in recent decades, been by far the most prevalent theory.

Phrase-Structure Grammar

Phrase-structure grammar was introduced by Noam Chomsky as a grammatical production system that categorized grammars and languages into a hierarchy of four types. A very similar notational scheme to describe one of those types (context free) was developed independently by John Backus for the purpose of describing programming languages. Phrase-structure grammar describes a language literally in terms of the structure of its phrases. The basic idea is that a sentence of a language is defined in terms of its phrase structure. The definition typically includes constituents that have their own phrase structure and hence require their own definitions. For example, in the sentence 'The big black dog chased the car', the phrase 'The big black dog' serves as the entity that is doing the chasing while the phrase 'the car' is the entity being chased. We could define the structure of a simple sentence as follows:

(1) S → NP VP
(2) NP → ART ADJ-SEQ NOUN
(3) VP → Verb | Verb NP
(4) ART → a | the
(5) ADJ-SEQ → ADJ | ADJ ADJ-SEQ
(6) ADJ → big | little | black | white | good | bad
(7) NOUN → boy | girl | dog | cat | car | truck
(8) Verb → chased | followed | saw

Each rule defines the phrase structure of one type of constituent (the arrow is read as 'is defined as'). The first rule defines a sentence (called an S) as a sequence of an NP followed by a VP. An NP is defined in rule (2) as being a sequence of three components, each of which requires further definitions. The first component,

which we have called ART, is defined by rule (4) and is either the word 'a' or the word 'the' (the vertical bar is read as a logical 'or'). The third rule requires a bit more explanation. The constituent ADJ-SEQ is defined, in part, in terms of itself. Such a definition is called a recursive definition. Using 'The big black dog' as an example, the ADJ-SEQ would be defined as an ADJ (the adjective 'big') followed by another ADJ-SEQ that consists only of an ADJ, which is the adjective 'black'. The process of matching of the actual words and phrases to the constituents defined by the grammar is called parsing, and the result of that matching is called a parse. Since, when one diagrams the matching, the result is a tree-like hierarchical structure, the result is often referred to as a parse tree.

Certain constituents, such as ART, ADJ, and NOUN, clearly indicate particular word classes. Generally, we assume a lexicon of the words of the language so that the grammar rules would assume a specific set of word classes:

(1) S → NP VP
(2) NP → article ADJ-SEQ noun
(3) VP → verb | verb NP
(4) ART → a | the
(5) ADJ-SEQ → adjective | adjective ADJ-SEQ

In such a characterization, an 'article' is any entry in the lexicon that is identified as belonging to the word class 'article'. Nouns, adjectives, verbs, and other classes are similarly defined.

Although the example is greatly oversimplified, it will be adequate to illustrate many of the major issues addressed by various grammatical models. Several points should be noted by this first simple example:

(1) Each rule defines a constituent, regardless of the context of that constituent, and is therefore called a context-free grammar. There are various ways to extend context-free grammars to account for context and some of those extensions are included in the grammatical models discussed below. Note also that phrase-structure grammars are not necessarily context free. In Chomsky's hierarchy, context-free grammars are in the second lowest (in terms of expressive power) level of the hierarchy. Generally, context-free grammars are highly desirable because of the relative ease in reading and understanding them. Greater expressive power may be achieved by allowing the left-hand side of a rule to contain more than one symbol, but such models are seldom used. Generally, other kinds of extensions are used.

(2) We omit discussions here of whether morphological rules are accounted for by the lexicon or the grammar. Hence, the word 'chased' may be in the lexicon explicitly, or there may be a morphological component that identifies 'chased' as being the past tense of the lexicon entry 'chase'. Although that is not relevant for our example, it is certainly relevant in general when evaluating grammatical theories. Some languages (e.g. Turkish, Hungarian, Finnish, or Mohawk) have morphological complexities that would make the positing of a lexicon that holds all morphological forms untenable.

(3) The discussion so far has addressed only the syntax of the language. It has not addressed semantic, pragmatic, or other issues.

(4) The example deals only with defining a sentence, but generally any model that attempts to deal with meaning must address larger units of language in order to account for context.

Categorial Grammars

The term categorial grammar was introduced by Yehoshua Bar-Hillel and combined some of Bar-Hillel's own work with earlier work of some Polish logicians, most notably Kazimierz Adjukiewicz. Categorial grammars in their original form are equivalent in expressive power to context-free phrase-structure grammars, but are based much more on words and their influence on adjacent or nearby words than are phrase-structure grammars. Each word class is a 'category'. Categories are either basic categories or derived categories. Derived categories have one or more operators in them that allow them to be combined with adjacent words in a manner much like multiplication. In the following example, the only basic category is C. Hence, a phrase 'the boy' would have the categories (NP/C)C. Treating the slash much as one would division, we can 'multiply' (NP/C) by C, arriving at NP. Hence, the phrase 'The boy' will be an NP. Similarly, the phrase 'The big boy' would have the category sequence (NP/C)(C/C)C, which, when 'multiplied' out, gives NP.

A categorial grammar equivalent to the phrase-structure grammar given above would have only one rule and a lexicon as follows:

Grammar Rule	Lexicon
(Ci/Cj)Cj → Ci	a:(NP/C)
	big: (C/C)
	black(C/C)
	boy:C
	cat:C
	dog:C
	girl:C
	little(C/C)
	the: (NP/C)
	white(C/C)

The example of categorial grammar has only one grammar rule. The rest of the language structure is in the lexicon. Although categorial grammars generally have more than one rule, the number of rules is always small (often two) with the structure primarily being handled by the lexicon. The underlying principle of categorial grammars is that each word is a functor, an object that operates as a function that maps one or more adjacent words into a new object, generally a new function.

Relational Grammar

Relational grammar is a direct descendant of transformational grammars that was developed primarily by David Perlmutter and Paul Postal in the early 1970s. One important characteristic of relational grammars is their cross-language generality. Early transformational grammars dealt almost entirely with English, but relational grammars were developed using a large variety of languages. Relational grammars view syntactic structures as a hierarchy of relations that are taken to be uninstantiated primitives. For example, the subject is level 1, the direct object is level 2, and the indirect object is level 3. Clausal structure is represented by a hierarchical diagram, called a stratal diagram, of the clause's terms.

One of the main motivations for the creation of relational grammars was to correct a perceived weakness of transformational grammars. Perlmutter and Postal argued that logical sentence components such as the subject are defined directly in terms of the sentence structure, which did not provide for a general characterization of relationships such as that between active and passive voice. As a simple example, 'The dog chased the boy' and 'The boy was chased by the dog' have different syntactic subjects, but the same logical 'subject'. That is, the dog does the chasing in both sentences. Hence, a model that defines the term 'subject' strictly in terms of constituent structure will have difficultly accounting for the similarity of the two sentences.

Case Grammar

Another descendant of transformational grammar is case grammar as introduced by Charles Fillmore in the early 1970s. Case grammar is based, according to Fillmore, on two principles: (1) the centrality of syntax and (2) covert categories. The former implies that the sentence's syntactic structure is primary (as opposed to, say, morphology) and the latter refers to the dominance of the meaning underlying the use of words and constituents. For example, in the sentence 'The boy hit the dog with a stick', syntactically, 'the boy' is the subject, 'the dog' is the direct object, and 'a stick' is the object of a preposition. A model consistent with case grammar would view the sentence as requiring the covert categories 'actor', 'affected', and 'instrument'. In the example, 'the boy'

would be the actor, 'the dog' would be the affected, and 'a stick' would be the instrument.

One place where we can see a more illustrative example is in the analysis of coordination. For example, given that Tom is cooking and Susan is cooking, it is reasonable to assert that Tom and Susan are cooking. However, given that Tom is cooking and the omelet is cooking, we would not say that Tom and the omelet are cooking (except, perhaps, to be humorous.) The point here is that although the sentence clauses 'Susan is cooking' and 'the omelet is cooking' appear to be syntactically similar sentences, the role that 'Susan' fills in the sentence is quite different from the role filled by 'the omelet'. In this example, the respective roles would be difficult to distinguish via purely syntactic means. Of course, sometimes the roles cannot be determined by syntactic means but depend on semantics and the context of the situation. For example, in a cartoon, Susan and a chicken could be cooking in the sense that both are in the act of preparing a meal, or both could be in the other role in a context involving cannibals.

Dependency Grammar

An alternative to describing the syntax of a language via its phrase structure is to describe the dependencies that exist among the words, phrases, and clauses. The idea is a very old one, but the modern characterization is based largely on the work of Lucien Tesnière in the late 1950s. The dependency approach to syntactic structure was overwhelmed by Noam Chomsky's transformational grammar approach and was thus largely ignored by the modern linguistic community. In recent years, more interest has been shown in the dependency approach as well as other alternatives to phrase-structure representations for syntactic structures.

Dependency grammars take the head and modifier approach to constituent structure. For example, in a phrase-structure approach, one might analyze the sentence 'cars burn gas' by focusing on the verb, 'burn'. Cars are the things doing the burning and gas is the stuff being burned. An alternative approach would be to consider cars as the head of the sentence and the burning of gas as an attribute of cars. Basically, the 'head' of a sentence, clause, or phrase is where we focus our attention. The dependencies are generally determined via the semantics of the sentence. An easy thumb rule is to imagine that you are recording the information of the sentence in, say, some kind of encyclopedia. The head of the phrase is then the entry under which the information is stored. In the case of 'cars burn gas' it is likely that storing the information under 'car' rather than 'burn' would ultimately prove to be more useful. Dependency grammars have often been used in information retrieval systems, particularly early systems.

Lexical-Functional Grammar

Lexical-functional grammar was developed in the early 1970s by Joan Bresnan and Ron Kaplan to provide a model that could serve as the basis of both a computationally feasible and psychologically realistic model of human language. Like relational grammar, it deviated from the transformation model by including nonstructural factors in the syntactic component of language. Also, like relational grammar, it gives a distinguished status to the notion of 'subject' and 'object'. The 'lexical' in the title is due to its emphasis on the role of the lexicon.

In a lexical-functional description of a sentence, there are two components: a constituent structure and a functional structure. The constituent structure is strictly structural and is based on a context-free phrase-structure grammar. The functional structure is by additional computations, expressed as equations, associated with the context-free rules. The computational aspects of lexical-functional are quite similar to a different formalism, developed in the late 1960s by Donald Knuth, called an attribute grammar. The basic idea is that each constituent of the language has a set of attributes. For example, the 'number' attribute of the NP 'the big black dog' could be ultimately assigned as 'singular' since the words 'the' and 'dog' are both marked as singular by the lexicon. Note that if their number attributes had been different, the noun phrase would not have been well formed.

Lexical–functional grammar is similar to case grammar in that the attributes constitute a set of named slots that are filled by the computational rules. Note also that attributes and their associated computational rules need not be confined to syntactic properties. Semantic features with complex values, including values that are actually active computational agents such as automata or computer programs, fit naturally into the model.

Generalized Phrase-Structure Grammar

Unlike lexical–functional grammar, generalized phrase-structure grammar (GPS) does not posit a separate, distinct functional component. The model presented here was introduced by Gerald Gazdar, in the late 1970s, but the formal mechanisms are very similar to W-grammar as developed by A. Van Vijngaarden in the late 1960s. The basic idea is to use phrase-structure grammar, but to allow more powerful *metarules* that define additional rules rather than syntactic constituents. This seemingly small change in the formalism changes the expressive power of the grammar substantially and allows the specification of complex syntactic rules such as matching a sentence in the active voice to its corresponding passive sentence. In particular, it allows GPS to dispense with the transformations of transformational grammar.

In Gazdar's model, grammatical constituents included components called derived categories, which were denoted using notation similar to the derived categories in categorial grammar. However, in GPS the derived category X/Y (where X and Y stand for grammatical constituents) represents a constituent X that is missing the constituent Y, that would normally be included as a component of X. This ability to define constituents with 'holes' allows a relatively straightforward treatment of sentences such as 'Which book is that article on Gazdar in?' In this example, the verb phrase of the sentence is missing a noun phrase ('which book'), and hence requires a noun phrase elsewhere in the sentence that can 'fill in the hole'.

Cognitive Grammar

Cognitive grammar was developed in the mid-1970s by Ronald Langacker. It was originally called space grammar and was considered by its creator to be fundamentally at odds with current linguistic trends of the time in that it claimed the inseparability of syntax and semantics, rejected the primacy of formal logic as the foundation of grammar, and argued for integration rather than specialization of theories for separate linguistic components.

Cognitive grammar, as its name implies, considers language as an integral part of human cognition and therefore knowledge of language cannot be wholly independent of knowledge of human cognition. Langacker argues that a separate processing unit that deals with reason and cognition has not been firmly established and may not exist. His model is intended to be valid whether such a separate module exists or not.

There are exactly three kinds of grammatical structures in cognitive grammar: semantic, phonological, and symbolic. Symbolic structures are not distinct from the other two, but combine them. Cognitive grammar attempts to characterize the mapping and providing the correspondence of the sounds we hear to the thoughts we associate with those sounds.

Cognitive grammar is based on several underlying principles. First, formal logic is not suitable for modeling the meaning of a sentence, since meaning is not based on truth values, but rather individual experiences. Second, the meaning of a particular word is defined by the universe of experience of the speaker (or listener) and not by a constrained formula such as is implied by a dictionary entry. Finally, like words, linguistic categories are not simple structures, but generally complex networks of linked structures.

Definite Clause Grammar

Definite clause grammar was first proposed by Alain Colmerauer in the late 1970s and developed further by Fernando Pereira and David Warren. Definite clause grammar is primarily a computational model that is based directly on logical formulas. For example, the phrase structure rule of the first example, S → NP VP, could be written as the logical formula:

For all sequences of words x, y, and z, if x is an NP and y is a VP and the concatenation of x and y is z, then z is an S.

All of the rules of any Phrase-structure grammar could be written according to such a logical formula. The purpose of stating a rule in that fashion is that it allows for direct implementation of a parser on a computer. Definite clause grammar was originally created as a natural way to write parsers in the programming language Prolog. Like GPS, definite clause grammar does not posit a separate distinct functional component, but simply adds additional components to the logical formula. For example:

For all sequences of words x, y, and z, if the concatenation of x and y is z and either (1) x is a singular NP and y is a singular VP, or (2) x is a plural NP and y is a plural VP, then z is an S.

This is a logical formula that would include number agreement between the noun phrase and the verb phrase. The logical formulas can be translated directly into computer code that provides a parser.

References

Blake, Barry J. 1990. Relational grammar. London: Routledge.

van Brenthem, Johan, and Alice ter Meulen (eds.) 1997. Handbook of logic and language. Cambridge, MA: MIT Press.

Walter J. Cook, S.J., 1989. Case grammar theory. Washington, DC: Georgetown University Press.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum, and Ivan Sag. 1985. Generalized phrase structure grammar. Cambridge, MA: Harvard University Press.

Grishman, Ralph. 1986. Computational linguistics, an introduction. Cambridge: Cambridge University Press.

Koerner, E.F.K., and R.E. Asher. 1995. Concise history of the language sciences from the Sumerians to the Cognitivists. Cambridge: Cambridge University Press.

Langacker, Ronald W. 1987. Foundations of cognitive grammar, Vol. I: theoretical prerequisites. Stanford, CA: Stanford University Press.

Mel'čuk, Igor A. 1988. Dependency syntax: theory and practice. Albany, NY: State University of New York Press.

Oehrle, Richard T., Emmon Bach, and Deirdre Wheeler. 1988. Categorial grammars and natural language structures. Boston. MA: D. Reidel Publishing Company.

Sells, Peter. 1987. Lectures on contemporary syntactic theories: an introduction to government-binding theory, generalized phrase structure grammar, and lexical–functional grammar. Menlo Park, CA: CSLI.

Winograd, Terry. 1983. Language as a cognitive process: syntax. Reading, MA: Addison-Wesley Publishing Company.

ALTON SANDERS

See also **Bar-Hillel, Yehoshua; Chomsky, Noam; Fillmore, Charles; Morphology; Morphological Typology**

Grammatical Function

The concept of 'grammatical function' or 'grammatical relation' comes from traditional grammars, which differentiate four grammatical functions. For example, in a sentence like *John gave this book to Peter at school*, it can be said that *John* functions as a 'subject' of the verb *to give*; *this book* is the verb's 'direct object' and *to Peter* plays the role of its 'indirect object'; finally, the locative *at school* is 'oblique'. These labels underlie a network of relations between the verb and its arguments.

In early studies of natural language, a sentence was identified with the logical proposition, which has two essential components: the substance (subject) and its property (predicate). The subject is supposed to be known to both speaker and hearer, while the predicate adds something new about the subject. Therefore, subjects usually have a definite reference, and precede the predicative part of an utterance (the verb with its objects). To put it differently, grammatical subjects ensure the coherence of a discourse that leads from old to new information. This perspective was developed by Prague School linguists of the 1920s and is followed by modern functional approaches to grammar.

From a semantic point of view, grammatical functions are usually associated with different situational roles, known as 'thematic roles'. For example, the subject is usually associated with the role of Agent (or Actor) (***John** writes …*), while such roles as Theme (*to give **a book***) and Goal (*to give it **to John***) are related to direct and indirect objects, respectively. Oblique arguments can express Instrumental and Locative thematic roles (*to be loved **by Mary**, to live **in London***). However, with psychological verbs the direct correspondence between thematic roles and grammatical functions is problematic. For example, the verb *to fear* assigns the Experiencer role to the subject and the Theme role to the object (*John* (subject) *fears this story* (object)), while the synonymous verb *to frighten* assigns the same roles, but in inverse order (*This story* (subject) *frightens John* (object)).

With respect to relevant word forms, the subject, rather than the object, induces verbal agreement in a finite clause. In ***He** knows them*, for example, the verb is marked with a final -*s* indicating third-person singular. Thus, the verb is said to 'agree' with the subject *he,* which is a third-person singular pronoun. Moreover, in so-called accusative languages, case-marking on nouns and pronouns can serve as an indicator of grammatical functions. Thus, in English we observe different pronominal forms for subject and object, as in *He saw him,* where *he* is the form for nominative case (subject) and *him* is accusative (object). Nevertheless, the correspondence between case-marking and grammatical functions is different in so-called ergative languages (including Inuit, Georgian, Basque, Hindi, and many others). In these languages, the subject of an intransitive verb (***I** run*) has the same marking as the object of a transitive verb—that is, the equivalent of the English sentence *The teacher has seen **me*** would be something like *The teacher **I** has seen* in Burushaski (a language isolate spoken in northern Pakistan).

Finally, some more complex grammatical properties also differentiate the subject from the object. For example, subjects (but not objects) can 'control' a missing element in a subordinate clause: thus, in *Paul helps Mary without asking for money,* it is Paul (subject), not Mary (object), who does not ask for money, i.e. the complete sentence would be *Paul helps Mary without Paul asking for money.* The subject of the subordinate clause can only be left out if it is identical to the subject of the main clause. Notice that in *Paul helps Mary without Mary asking for money* the subject of the subordinate clause cannot be left out. Moreover, in some languages (e.g. Latin, Russian, Norwegian, Icelandic) there are reflexive pronouns whose antecedant must be the subject (so-called 'subject-oriented reflexives'). Accordingly, in the Icelandic sentence *Pall gaf barninu bók sín* (literally *Paul gave child book self's*), the book is construed as belonging to Paul (subject), not to the child (object).

In languages with rich noun inflection, a noun may behave syntactically as a subject but be marked as an object (so-called 'grammatical function-splitting'). Icelandic, for instance, has such constructions as *Mér likuði konan sín* (literally *Me likes wife self's*), *I like my wife*. Here, *mér* does not have the word form normally

associated with the subject, since it is dative (not nominative) and does not trigger verbal agreement. However, *mér* is the subject from the syntactic point of view, since it can refer to the same entity as a subject-oriented possessive *sín*. In the literature, such cases are known as the 'quirky subject' phenomenon.

Many linguists (following Noam Chomsky) maintain that objects are more closely related to the verb, and consequently more likely to form idioms than subjects (*John is **pulling my leg**, this house **catches the eye***, etc.). This asymmetry is said to be due to the fact that verbs and objects form the predicative unit, while the subject is a more independent entity. From this point of view, the notions of 'subject' and 'object' can be defined on structural grounds.

However, nonconfigurational languages, in which word order is free (e.g. the Australian language Warlpiri), have been used to argue against a purely structural interpretation of grammatical functions. Since the mid-1970s, some new theoretical approaches have been created on the basis of grammatical functions being autonomous categories. For example, grammatical functions may be regarded as syntactic primitives forming the following hierarchy: subject > direct object > indirect object > oblique. This is the stance taken by the framework of 'relational grammar'. It is argued that certain sentence structures may be described in terms of 'grammatical function-changing processes' . For example, the formation of passives, such as ***John has broken the glass—The glass has been broken (by John)***, is characterized as follows: the object is 'promoted' (in hierarchy) to the subject, while the subject is 'demoted' to the oblique (the presence of the latter is optional). Dative shift in English (e.g. *I gave my favorite book **to Mary*** and *I gave **Mary** my favorite book*) also represents 'grammatical function-changing', here the promotion of the indirect object to the direct object.

Another possibility is to conceive of grammatical functions as categories that relate thematic roles (Agent, Patient, Experiencer, etc.) to the concrete means of expression in a given language. For instance, the English lexicon has the verb *to invite,* whose Agent role is mapped to the subject, while its Patient role is mapped to the object, as in *John* (subject-Agent) ***invited** the president* (object-Patient). The passive counterpart of this verb is formed by remapping Patient to the subject and Agent to the oblique, as in *John* (subject-Patient) *is **invited** by the president* (oblique-Agent). This is the view elaborated in the framework of 'lexical-functional grammar.'

Summing up, the traditional notion of grammatical function is controversial in modern linguistics. The core grammatical functions, subject and object, cannot be defined absolutely, since they are differentiated on the basis of several possibly independent properties. From the semantic point of view, they represent the two main participants in the event denoted by the verb. Grammatical functions may be useful descriptively, since they make it possible to abstract away from language-specific properties and to generalize over certain syntactic phenomena.

References

Bhatt, Darbhe N. S. 1991. Grammatical relations: the evidence against their necessity and universality. London and New York: Routledge.

Bresnan, Joan. 2001. Lexical—functional syntax. Oxford and Malden, MA: Blackwell.

Davies, William D., and Stanely Dubinsky (eds.) 2001. Objects and other subjects: grammatical functions, functional categories, and configurationality. Dordrecht and Boston: Kluwer.

Perlmutter, David M. (ed.) 1983. Studies in relational grammar, Vol. 1. Chicago: The University of Chicago Press.

Ura, Hiroyuki. 2000. Checking theory and grammatical functions in universal grammar. New York and Oxford: Oxford University Press.

Van Valin. Jr. Robert D. and Randy J. LaPolla. 1997. Syntax: structure, meaning and function, Cambridge: Cambridge University Press.

EGOR TSEDRYK

See also **Case; Configurationality; Predication; Thematic Structure; Word Order**

Grammaticalization

Grammaticalization is defined as the development from words with full meaning to forms with only grammatical function, and from grammatical to even more grammatical forms. Since the development of grammatical forms is not independent of the grammatical constructions to which they belong, the study of grammaticalization is also concerned with constructions, and with even larger discourse segments. The primary goal of grammaticalization theory is to describe how grammatical forms and constructions

arise and develop through space and time, and to explain why they are structured the way they are.

In the history of grammaticalization studies, three main phases can be distinguished. The first phase is associated with the work of eighteenth-century French and British philosophers who argued that grammatical complexity and abstract vocabulary derive historically from concrete words. In the nineteenth century, it was mostly German linguists, especially Franz Bopp and Georg von der Gabelentz, who used findings of grammaticalization to understand grammatical change and language structure. The third phase started in the 1970s and was shaped by scholars using functionalist frameworks of linguistic analysis, most of all Talmy Givón. His slogan 'Today's morphology is yesterday's syntax' opened a new perspective for understanding grammar, and he proposed the following unidirectional cycle (where the endpoint marks the beginning of a new cycle again leading from Discourse to Zero; see Givón 1971, 1979):

Discourse > Syntax > Morphology > Morphophonemics > Zero.

While earlier work dealt mostly with the study of individual words and grammatical items, modern work is more concerned with the use of linguistic forms in constructions and larger discourse units.

The following are the main assumptions underlying grammaticalization theory. First, it is argued that language is a historical product and has to be accounted for, first of all, with reference to the historical forces that are responsible for its present structure. Accordingly, it is claimed that this theory offers more powerful explanations of language structure than any approach that is confined to the analysis of a language at a particular point in time.

Second, the development of grammatical forms is unidirectional, being the result of a process whereby forms for concrete meanings are used to express more abstract (grammatical) meanings as well, e.g. when a form used for a visible object (e.g. the body part *back*) is also used to refer to a nonvisible item (the spatial notion *behind*), or a form used for an action (*go to*) is used to also refer to a grammatical notion (FUTURE TENSE). But equally common are processes whereby items that are already part of the inventory of grammatical forms give rise to more strongly grammaticalized and more abstract items. A number of examples contradicting the unidirectionality principle have been found, but as acknowledged by most scholars who have identified exceptional cases, such examples are few compared to the large number of cases that conform to the principle (see especially Newmeyer 1998).

On account of its specific directionality, grammaticalization has been described in terms of metaphorical transfer, leading from the domain of concrete objects to that of space, from space to time, from ('real-world') space to discourse space, and so on.

There is wide agreement that grammaticalization involves three interrelated mechanisms: (1) 'desemanticization' (or 'semantic bleaching'), loss of meaning content; (2) 'decategorialization', loss of morphosyntactic properties (e.g. inflection) characteristic of lexical or other less grammaticalized forms; and (3) 'erosion' (or 'phonetic reduction'), loss of phonetic substance. Desemanticization results from the use of forms for concrete meanings that are reinterpreted in specific contexts as more abstract, grammatical meanings. Having acquired grammatical meanings, these forms tend to become increasingly divergent from their old uses: they lose categorial properties characteristic of their old uses, hence undergo decategorialization, and they tend to be used more frequently, become more predictable in their occurrence and, consequently, tend to lose sounds. While all three mechanisms involve a loss of properties, there are also gains. In the same way as linguistic items undergoing grammaticalization lose in semantic, morphosyntactic, and phonetic substance, they also gain properties characteristic of their uses in new contexts. Grammaticalization requires specific contexts to take place, and it can be, and has been, described as a product of context-induced reinterpretation or inferencing.

Grammaticalization thus begins with concrete, lexical forms and constructions, and ideally ends in zero, that is, grammatical forms increasingly lose semantic and phonetic content and, in the end, they may be replaced by altogether new forms; grammaticalization has therefore been described as a cyclical process. Cyclicity is a frequent but not a necessary property of grammaticalization processes. Furthermore, grammaticalization has been described by some as a process that involves the reanalysis of grammatical forms and/or constructions, while other authors argue that there is no necessary relationship between grammaticalization and reanalysis.

References

Bybee, Joan L., Revere D. Perkins, and William Pagliuca. 1994. The evolution of grammar: tense, aspect and modality in the languages of the world. Chicago: University of Chicago Press.

Givón, Talmy. 1971. Historical syntax and synchronic morphology: an archaeologist's field trip. Chicago Linguistic Society 7. 394–415.

————. 1979. On understanding grammar. New York, San Francisco, London: Academic Press.

Haspelmath, Martin. 1998. Does grammaticalization need reanalysis? Studies in Language 22(2). 315–51.

Heine, Bernd. 1997. Cognitive foundations of grammar. Oxford, New York: Oxford University Press.

Heine, Bernd, Ulrike Claudi, and Friederike Hünnemeyer. 1991. Grammaticalization: a conceptual framework. Chicago: University of Chicago Press.

Heine, Bernd, and Tania Kuteva. 2002. World lexicon of grammaticalization. Cambridge: Cambridge University Press.

Heine, Bernd, and Mechthild Reh. 1984. Grammaticalization and reanalysis in African languages. Hamburg: Buske.

Hopper, Paul J. 1991. On some principles of grammaticization. In Traugott and Heine, pp. 17–35.

Hopper, Paul J, and Elizabeth C. Traugott. 1993. Grammaticalization. Cambridge: Cambridge University Press.

Lehmann, Christian. 1985. Grammaticalization: synchronic variation and diachronic change. Lingua e Stile 20(3). 303–18.

Newmeyer, Frederick J. 1998. Language form and language function. Cambridge, MA: MIT Press.

Traugott, Elizabeth C., and Bernd Heine (eds.) 1991. Approaches to grammaticalization, Vol. 1. Amsterdam, Philadelphia: Benjamins.

BERND HEINE

See also **Bopp, Franz; Givón, Talmy**

Great Britain

Great Britain, or more properly the United Kingdom of Great Britain and Northern Ireland, includes England, Scotland, and Wales and Northern Ireland. The total area of the current United Kingdom is 244,820 square kilometers with a total population of 59,511,464 (2000 est.).

Wales was politically integrated with England in 1536, Scotland officially joined England and Wales in 1707 with the Act of Union, and Ireland was added in 1801 to create the political union of the United Kingdom of Great Britain and Ireland. The union with Ireland was in effect until the Irish Free State was formed in 1921. The six counties of Northern Ireland, with a predominantly Protestant population, remained in the union, and the official name was changed to the United Kingdom of Great Britain and Northern Ireland in 1927. Now known as the Republic of Ireland (Eire), the Irish Free State became a separate republic in 1949.

The approximate historical periods of English within Great Britain are recognized as Old English from 450 to 1100, Middle English from 1100 to 1500, early modern English from 1500 to 1700, and modern English from 1700 to the present.

Many languages have existed in Great Britain; however, English has been the dominant language since the sixteenth century, with various other languages and dialects existing throughout the area. Languages that were once in Great Britain but are now extinct include Anglo-Danish, Anglo-Latin, Anglo-Saxon (Old English), Brythonic, Cornish, Manx Gaelic, Norman French, Norn, and Pictish.

Indigenous languages currently found in Great Britain include English, Irish Gaelic, Scots, Scottish Gaelic, and Welsh. Various other immigrant languages (including Caribbean English Creole, Hindi, Urdu, Punjabi, and Cantonese) are now spoken in cities like Birmingham, Cardiff, Glasgow, Leeds, and London.

England

England, which gets its name from the Old English *Englaland* (the land of the Angles), is in the southern part of the island of Britain and is the original home of English, a Germanic language that came with the Angles and Saxons to England from modern-day northern Germany, starting in the fifth century. The Celts, the earlier inhabitants, intermarried with the Anglo-Saxons or migrated to the outer parts of the British Isles. Even though the Anglo-Saxons lost their power in 1066 with the Norman Conquest, English and England remained strong, and by the sixteenth century, English was the official language of England.

Various minority groups live in England, using languages other than English as their only language or as a bilingual partner to English. The Irish are the largest minority group, and there are also groups of Europeans (Italians, Greeks, Turkish Cypriots, and Poles), Asians, Africans, Arabs, West Indians, Pakistanis, and Indians. In addition, Welsh and Cornish are also still spoken. Cornish, a Celtic language, has been revived and is now taught in some Cornwall schools, and some speakers of Welsh in places like London maintain their language.

English

English is the official language and is sometimes referred to as Anglo-English, England English, English English, and British English. The term 'British English' is often used when trying to compare the dialects of England with dialects around the world.

Many social and regional English dialects exist. Even though some people from different dialect groups may sometimes have difficulty understanding each other, the media has helped spread standardized words and pronunciations.

Standard English and Received Pronunciation

Standard English (SE) and Received Pronunciation (RP) are prestige social forms of English. SE, in its linguistics sense, relates to mostly the standardized grammar and vocabulary of English, and RP to the pronunciation of the educated upper-middle and upper class. However, the terms are often used interchangeably by those outside of linguistics. Even though it is assumed that SE includes RP, various accents are part of SE use. Most people write SE with slight variations in grammar and spelling; many also speak SE with some variation in regional vocabulary. Some speakers use two dialects—a social or regional one for friends and family and a version of SE for formal use, and most urban and rural areas have a variety of spoken English that has been influenced in some way by social or regional factors.

Cockney

Another social dialect at the opposite end of the social spectrum from RP is Cockney, the speech of about 7 million working-class Londoners. Although often stigmatized, Cockney has a long history and makes up the most prominent and widely spoken urban dialect in Great Britain.

Regional Dialects

Different regional dialects of English can be traced back in England to the eighth century when the Germanic tribes, who had settled in different areas of England in the preceding three centuries, had *Englisc* as their common language. The Angles settled in the Midlands and along the eastern coast, and their dialects included Mercian and Northumbrian. The Jutes settled around Kent and spoke the Kentish dialect. The Saxons settled in Essex, Middlesex, Sussex, and Wessex. Under the leadership of Alfred the Great, the West Saxon dialect of Wessex became the dominant dialect and served as the prestige and literary dialect of the Middle Ages.

Eleventh-century dialects included Northern, West Midland, East Midland, Southern, and Kentish. After the Norman Conquest, the East Midland dialect from the London and East Anglian region became the new standard with the help of William Caxton, the first major British printer, who used East Midland as the printing standard. Regional speech came to be considered as substandard but was used well into the eighteenth century by the middle and upper classes.

Today, regional varieties of English dialects are still strong, especially in large cities and rural areas. Regional dialects include South-East English around the London area, South-West English around the Cornwall area, West Midlands English around the Birmingham area, East Midlands English around the East Anglian region, North-West English around Liverpool and Manchester, and North-East English around Leeds, York, and the Yorkshire countryside.

Wales

Wales, in the southwest corner of the island of Britain, gets its name from the Old English word *Wealas* (foreigners). Celts from northwestern Europe originally settled in this area and were subsequently invaded by Romans. When Anglo-Saxons took control of England in the fifth century, Wales was the only Celtic region left in southern Britain, and the Welsh fought off the English, Irish, Scandinavians, and Normans from the fifth to eleventh centuries. In the thirteenth century, the last native Welsh prince was killed during an English invasion, and from 1301 on, the Prince of Wales has been British. In 1536, Wales was officially united with England, and English became the official language.

Welsh English

Close to 100% of the Wales population speaks English. Welsh English represents a spectrum of varieties and includes the social dialects of bilingual Welsh/English speakers, the English of the working class, SE speakers with a Welsh accent, and SE speakers with an RP accent. Welsh has a strong influence on the northern, mid, and southern regional dialects of Wales, with a very strong influence in the northern counties, where English/Welsh bilingualism is common.

There is little or no stigma attached to the use of Welsh English in Wales; however, many believe that education should be bilingual and hence all Welsh people have access to Welsh as the national language. Others, especially non-Welsh speakers, do not believe that speaking Welsh is a necessary part of being Welsh. Tensions are high, however, concerning the use of the Welsh language, since many believe that education should be bilingual so all Welsh people have access to Welsh as their national language.

Welsh

Welsh, the Celtic language of Wales, is known as *Cymraeg* to its speakers and was the main language of Wales for many centuries. However, Welsh use has been diminishing since the fifteenth century. An eighteenth- and nineteenth-century revival led to Welsh being taught in all schools; in some schools, it is even the language of instruction. Currently, about 26% of the population is bilingual (with Welsh and English); however, only a few monolingual Welsh speakers exist. Even so, Welsh is relatively stable; a

Welsh-only television channel and bilingual road signs show that its steady decline might be slowing. Welsh has several dialects that differ mostly in pronunciation; the vocabulary has mostly been standardized by education and the media.

Scotland

Scotland, a part of Great Britain since 1707, gets its name from the Old English 'land of the Scots' and includes the mainland, which is divided into three main regions (the Highlands, the Lowlands, and the Southern Uplands), and the outlying islands, which include the Western Isles (also known as the Hebrides) and the Northern Isles (the islands of Orkney and Shetland).

When the Romans invaded Britain in the first century, they called the area now known as modern Scotland by the name *Caledonia* and called its inhabitants the Picts (from Latin *picius* 'painted') because of the Pict tradition of painting their bodies. In the fifth century, after the Romans withdrew from England, and Germanic tribes began their invasions into England, Christian Celtic immigrants from Ireland began to settle the area of modern Scotland, converted the Picts to Christianity during the sixth century, and added the Pictish kingdom to the Scots kingdom by the ninth century.

By the tenth century, this area was known as Scotland. After the Norman Conquest, many Anglo-Saxons settled in the Lowlands of Scotland, bringing English customs and the English language. England and Scotland were joined formally together in the Act of Union of 1707 in which both parliaments were merged, thus creating the Kingdom of Great Britain.

Scotland's language history is complex. In the Middle Ages, six languages were in regular use: Cumbric, Gaelic, Inglis, Norn, Norman French, and Latin. Modern Scotland has two or three (depending on who does the analysis) languages: Scottish Gaelic, English (Scottish and RP), and Scots. Up to the fifteenth century, Scots was the term applied only to Gaelic and its speakers; now this variety is most commonly referred to as Gaelic, Scots Gaelic, or Scottish Gaelic. From the late fifteenth century, the term Scots has referred to Scots English, the language of the Lowlands. Scottish English refers to the other varieties of English used in Scotland and, depending on the analysis and the discussion, can include or exclude Scots English.

Scottish Gaelic
Gaelic is the English name for the Celtic language of Ireland, Scotland, and the Isle of Man. In Scotland, it was formerly called *Erse* or *Irish,* but speakers today often refer to it as *the Gaelic*.

Until the eighteenth century, Gaelic was the principal language of Ireland; however, with the increased influence and pressure of Northern English, the use and prestige of Gaelic have declined since the twelfth century. Since the late twentieth century, Gaelic has been a first language for very few people (usually in the Hebrides) but has regained some of its prestige through efforts to encourage bilingual policies, and has made Gaelic a part of the educational system. According to the 1981 census, approximately 80,000 people speak Gaelic, but most of these are bilingual English/Gaelic speakers.

Scottish Gaelic has a strong literary tradition, and there is a resurging twentieth-century Gaelic literature encouraged by *An Comunn Gaidhealach* (The Highland Association) and *Comunn na Gàidhlig* (The Gaelic Association), two organizations that support and promote the Gaelic culture and language.

Scottish Gaelic has influenced Scots English and other varieties of Scottish English in pronunciation, syntax, and vocabulary, and this influence runs both ways.

Scots
Some language scholars consider Scots (or Scots English) a language and others consider it a dialect of English. Whatever way it is defined, since the beginning of the eighteenth century, scholars have specialized in studying it in its different periods: Old English (to 1100), Early Scots (1100–1450), Middle Scots (1450–1700), and Modern Scots (1700 onward).

Scots is first recorded in the seventh century in southern Scotland as a variant of the Old English of Northern England. By the fourteenth century, this variety known as *Inglis* had replaced Gaelic as the language of the Lowlands. By the mid-sixteenth century, Scots began to be heavily influenced by the English of southern England. In the eighteenth century, Scots was considered substandard, but a revival of vernacular literature in Scots in the nineteenth century encouraged the middle and upper classes to become more interested in the Scots language. In the twentieth century, there have been more attempts to revive Scots as the national language; even so, English and Scottish English are still the dialects of power and prestige, and there has been a steady decline of native Scots speakers.

The dialects of Scots fall into four main regional dialect areas: the Northern Isles Scots of Orkney and Shetland, the Northern Scots of northern Scotland, the Southern Scots near the English border, and the Central Scots of middle Scotland, including the working-class dialects of Edinburgh and Glasgow, which are often referred to pejoratively as Gutter Scots.

Scottish English

If Scots is a language, then Scottish English includes all other varieties of English in Scotland, including the Highland English of the Highlands and Islands, the Scottish standard English of the past three centuries, and the RP dialect of the upper middle and upper class. Scottish English has many similarities with the English of England but also has many features of Scots, and many modern speakers of Scottish English acquired Scots as their first language in their childhood.

Northern Ireland

Northern Ireland, a part of the United Kingdom, includes six of the nine counties that originally made up the early Celtic kingdom of Ulster: Antrim, Armagh, Derry or Londonderry, Down, Fermanagh, and Tyrone. The Celt leaders, who were Druids, lost their power in the fifth century after St. Patrick introduced Christianity. After many centuries of Celtic challenges to both Christianity and the power of England, the Celt leaders fled Northern Ireland in the early seventeenth century, and about two thirds of the current Northern Ireland population are descendants of Protestant English and Scottish settlers who came to Ulster in the same century. The other third are Irish in origin and Catholic. By the mid-nineteenth century, English was the language of power and prestige. In modern Northern Ireland, an omnipresent tension exists between Protestants and Catholics, and even though the right to free language choice is part of this tension, both sides of the argument are likely to use some variety of English now.

Irish Gaelic

Irish Gaelic, or Irish, is a Celtic language that remained the main language of Ireland until the eighteenth century when the British passed and enforced laws that made Irish the language of the powerless. The school system was restructured with new National Schools, in which English was the only language of instruction. Gaelic was also negatively impacted by the Irish famine of the mid-nineteenth century, which resulted in the mass emigration of Gaelic first-language speakers.

By the end of the nineteenth century, there was a revival of interest in the Irish language, literature, and history of Northern Ireland and the Republic of Ireland. This new national attitude helped slow the steady decline of Irish Gaelic but only after English had already gained enormous power and Irish had died out as a spoken language except in isolated rural areas of Northern Ireland.

Northern Irish English

Northern Irish English refers to the English used in Northern Ireland; however, there are four distinct varieties: Ulster Scots, Anglo-Irish, Hiberno-English, and Belfast English.

Ulster Scots

Ulster Scots, also known as Scotch-Irish, was brought to Ireland in the seventeenth century by about 200,000 migrating Lowland Scots and is mainly found in parts of Antrim, Derry, and Down counties in Northern Island. Ulster Scots is usually identified as the dialect of the working class and is found throughout Northern Ireland and the Irish Republic.

Anglo-Irish

Anglo-Irish, or Ulster English, is a variety of English spoken over most of Northern Ireland and the Irish Republic, except for the most northern counties. Like Ulster Scots, the beginnings of Anglo-Irish came to Ireland in the seventeenth century, but with English settlers and not Scottish. There are regional and social varieties that are heavily influenced by the education of the speaker. RP is usually part of the Anglo-Irish of the middle class, but less educated speakers usually have other distinct pronunciations. As time goes by, a standardized variety of Anglo-Irish is becoming more common.

Hiberno-English

Hiberno-English is another variety of English in Ireland, used mainly by less educated speakers. Mostly found in rural counties (Armagh, Fermanagh, and Tyrone) in Northern Ireland, it is heavily influenced by Irish Gaelic.

Belfast English

Belfast English, another variety of English in Northern Ireland, varies with the level of education of the speaker. It is normally described as rapid informal speech that has the vocabulary of Ulster Scots with some nonstandard grammatical features of Anglo-Irish.

References

Baugh, Albert C., and Thomas Cable. 1992. A history of the English language. Englewood Cliffs, NJ: Prentice-Hall.
Crystal, David. 2000. Language death. Cambridge: Cambridge University Press.
Hughes, Arthur, and Peter Trudgill. 1998. English accents and dialects: an introduction to social and regional varieties of English in the British Isles. London: Edward Arnold.
McArthur, Tom. 1998. The English languages. Cambridge: Cambridge University Press.
Price, Glanville. 2000. Languages in Britain and Ireland. Oxford and Malden, MA: Blackwell Publishers.
Wells, John C. 1982. Accents on English: an introduction. Cambridge: Cambridge University Press.

ALLISON SMITH

See also **Celtic Languages**

Great Vowel Shift

English is unique among the European languages using the Roman alphabet in that it has a special set of 'names' for the vowel letters <i, e, a, o>. The names reflect the pronounciation of these letters when they represent long or tense vowels, as in *mine, cede, fame, grove*. Until the fourteenth century the value of these letters was close enough to their Latin pronounciation so that Chaucer (d. 1400) could rhyme *A* with *omnia*:

> On which was first i-write a crowned A,

> And after, *Amor vincit omnia. Prol.* 161

The long vowels acquired their English names subsequent to a set of changes known collectively as *The Great Vowel Shift* (GVS). The GVS is one of the main reasons for the apparent anomalies of English spelling today. The changes covered by this label are universally acknowledged as the most dramatic readjustment of the system of long vowels in the recorded history of English. The phenomena subsumed under the general designation of GVS have been the focus of extensive scholarly inquiry during the last 150 years, and the debate is ongoing.

The research history of the GVS encapsulates the history of phonological research in general, and the history of English historical language studies: from atomistic description, to structuralist and generative generalizations, and, more recently, increased attention to surface phonetic facts in the context of regional and social variation. Nineteenth-century philologists described the changes in great detail, focusing on the results in southern standard British English, or, more specifically, on Received Pronounciation (RP). The latter is the supraregional variety of British English based on nineteenth-century southern pronounciations, and fostered until at least the 1960s as the standard in the media and the 'correct' pronounciation in (British) English instruction around the world. The vowel correspondences reconstructed for the changes between late Middle English (ME), c. 1350–1400, and twentieth-century RP are as follows (see Figure 1).

The last century has witnessed both the 'making' and the 'undoing' of the GVS. Starting with Karl Luick (1898) and Otto Jespersen (1909), the changes shown above have been treated as a unified set of phonological events, where each new value is supposed to be the result and the trigger of a chain-like shift of all ME pure long vowels. In most cases, the phonemic contrasts between the original entities were preserved, although their phonetic realizations were changed. This position is well represented by Jespersen's summary statement (1909:231): 'The great vowel-shift consists in a general raising of all long vowels with the exception of the two high vowels /i:/ and /u:/, which could not be raised further without becoming consonants and which were diphthongized into …[ai, au].' The presumed structural connectedness among the changes is represented in the chart below:

aɪ ← i: u: → aʊ
      ↑   ↑
      e: o:
      ↑   ↑
      ɛ: ɔ:
      ↑
      a:
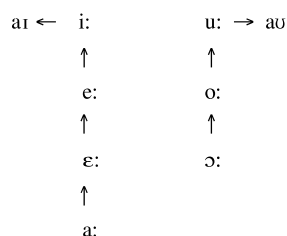
The all-encompassing, chain-shift view of the GVS dominated the scholarship throughout the early parts of the twentieth century. Hypotheses addressing the initiation, causation, and propagation of the putative massive chain shift have been proposed in terms of numerous phonological theories, including structuralism, generative phonology, lexical phonology, dependency phonology, particle phonology, lexical diffusion, and optimality theory. With rare exceptions, a representation of the changes from Middle to Modern English, with each long vowel linked to an arrow pointing upward or outward, is still repeated in textbook accounts of the history of English phonology. The characteristics associated with the position that the GVS was a comprehensive and coherent chain shift of the ME

c. 1350-1400	RP	Examples
i: ———>	aɪ	*bind, design, wife*
e: ———>	i:	*fiend, see, tree*
ɛ: ———>	i:/eɪ	*meal, sea, steak*
a: ———>	eɪ	*bake, save, strange*
u: ———>	aʊ	*cow, crown, how*
o: ———>	u:	*do, moon, scoop*
ɔ: ———>	əʊ	*boat, nose, stone*

Figure 1

monophthongal long vowels, and the reasons why this position has been challenged, are as follows:

(1) *Scope*. The GVS affected all and only long vowels.

It is true that all long vowels underwent some kind of change during the early Modern English period (c. 1450–1750). Nevertheless, any GVS representation ignoring the other functionally long vocalic units in ME is too restrictive. Diphthongal variants of the high vowel /iː/, /iy/, existed or arose in ME from earlier native and borrowed sequences of high vowels and following consonantal /j/ and /h/, thus ME /stiy/ 'sty', /sliy/ 'sly' from earlier <stig> /stij/, <sligh> /slij(h)/. The high back long vowel /uː/ had noncontrastive variants /uw/, as in ME /fuwl/ 'fowl', /buw(h)/ 'bough' from earlier native and borrowed <fugol> [fuɣ(ə)l], <boh> [buw(h)]. In both cases, the diphthongal and the steady-state variants merged toward a diphthongal realization. These initial diphthongs were further optimized by differentiation of their endpoints: /aɪ/ for *bind, design, wife, sty*, and /aʊ/ for *cow, crown, how, fowl*. Thus, /aɪ/ and /aʊ/ are the attested RP results of the gradual change of both 'pure' long vowels and minimal diphthongs of high vowels followed by a homorganic glide; it was the diphthongal realizations that were formative in terms of the later history of these vocalic units. Similarly, the history of the ME mid long vowels /ɛː/ (as in *steak*), /aː/ (as in *bake*), and /ɔː/ (as in *boat*) cannot be separated from the history of preexisting diphthongs as in *day, play, weigh, blow, stow, dough* from earlier /dæj, plej, wejh/, /blow, stow, dowh/. Therefore, any description of the reorganization of the vowel system of ME should take into consideration both the history of the long vowels proper and the diphthongal entities with which the long vowels merged in the course of the GVS.

(2) *Dating*. The GVS began by c. 1400, and was completed by c. 1700.

This chronological span can accommodate only some of the changes that shaped the long vowel system of Modern English. The more dramatic changes associated with the GVS, the full diphthongization of the high vowels /iː/ and /uː/ and the raising of the high mid vowels /eː/ and /oː/ in RP, were indeed arguably accomplished within this time-frame. However, the mergers of the 'pure' high vowels with preexisting diphthongs that determined the future of these vowel types were already under way in early ME, if not earlier. High-vowel diphthongization and mid-vowel raising can be traced in the spelling starting from c. 1350.

Another problem with such restrictive dating comes from the post-1700 instances of continuing shifts in non-RP varieties of English. Figure 2 shows some examples of post-GVS changes of the high vowels that are not covered by the traditional conceptualization of the shift as a single event bounded by cut-off time points. The first column represents the ME input, the second column shows the reflexes of these high vowels in RP, the third one shows the developments in London English, also Australian and New Zealand English, and the fourth column represents current realizations of these vowels in some southern US varieties of English.

The changes of the long vowels outside RP cannot be fitted within the conventional GVS picture. Regarding the GVS as a one-time coherent event, which began and ended at well-defined points of time, causes us to lose sight of the rich phonetic variation that precedes and follows it. Reference to the dating of the GVS must be accompanied by the realization that it is an artificially isolated portion of a continuum, representing the evolution of the entire phonological system of the language through time, with no starting- and no endpoints.

(3) *Mechanism*. The GVS started with the high and the high mid vowels.

(a) The high vowels /iː/ and /uː/ were diphthongized first. The raising of the high midvowels /eː/ and /oː/ was motivated by the existence of vacant slots for /iː/ and /uː/. This is the 'drag chain' theory that originates with Jespersen.

(b) The impulse for the diphthongization of the high vowels /iː/ and /uː/ came from the raising of the high mid vowels /eː/ and /oː/. This is the 'push chain' theory that originates with Luick.

The 'drag chain' position is now mostly of historical interest. Luick's observation that there was no diphthongization of /uː/ in those dialects of northern ME where /oː/ was fronted to /øː/ c. 1300, has been accepted as a good argument in favor of the causal link between the initial raising of the mid-vowels and the diphthongization of /iː/ and /uː/ in the southern dialects. In that sense, the chain linking the high and the mid high vowels is a 'push chain'.

ME	RP	Lon, Aus, NZ	Am. South	
iː~iy	aɪ	ɒi	aː	*bind, wife*
uː~uw	aʊ	æʊ	aː	*cow, crown*

Figure 2

(4) *Stages*: The GVS proceeded in two distinct unrelated stages.

Separating the developments of the long vowels into two distinct stages is a relatively recent addition to the GVS accounts. Earlier, GVS scholarship had assumed that all long vowels were proper participants in a single 'great' chain shift, yet research during the last three decades has shown that only the top two heights of the ME long vowel system were involved in chain shifting. The low mid vowels of southern ME, /ɛ:, ɔ:/ and the low vowel /a:/ began to raise and merge with preexisting /e:~ey/, /o:~ow/ only in the late sixteenth century. The merger of the vowels in e.g. *see* (ME /e:/> /i:/ c. 1550) with the vowel of *sea* (ME /ɛ:/>/e:/ c. 1650) did not begin until the end of the seventeenth century, and continued into the eighteenth century. These later developments are no longer claimed to be part of the chain shift that can properly be reconstructed for the four top vowels in southern standard British English.

In summary, the term GVS is not well defined. It has been applied as a cover term for a variety of often unrelated changes of the southern English long vowels and diphthongs that started in ME and can be shown to continue to this day. In that sense, it includes raisings, diphthongizations, mergers, and minichain shifts within the long vowel system of one language variety. A redefinition of the term GVS, suggested by Lass (1992:153), restricts it to the changes of the original high and high mid vowels [i:, u:, e:, o:] occurring in southern English in the fifteenth to sixteenth centuries. Since the remaining vowels do not participate in chain shifting in a coherent way, and only a minichain shift can be reconstructed reliably, the survival of the term Great Vowel Shift as a synonym for a comprehensive chain shift in the history of all English long vowels is in doubt.

References

Jespersen, Otto. 1909. A modern English grammar on historical principles, Vol. 1. Heidelberg: Carl Winter's Universitätsbuchhandlung.

Lass, Roger. 1992. What, if anything, was the Great Vowel Shift? History of Englishes: new methods and interpretations in historical linguistics, ed. by Matti Rissanen, Ossi Ihalainen et al. Berlin: Mouton de Gruyter.

———. 1999. Phonology and morphology. The Cambridge history of the English language, Vol. III, 1476 to 1776, ed. by Roger Lass. Cambridge: Cambridge University Press.

Luick, Karl. 1898. Untersuchungen zur englischen Lautgeschichte. Strassburg: Trübner.

———. 1921–1940. Historische Grammatik der englischen Sprache. Leipzig: Tauchnitz.

Stockwell, Robert P. 1972. Problems in the interpretation of the great English vowel shift. Studies in linguistics: papers in honor of George L. Trager, ed. by M. Estellie Smith, 330–53. The Hague: Mouton.

———. 1978. Perseverance in the English vowel shift. Recent developments in historical phonology, ed. by Jacek Fisiak. The Hague: Mouton.

Stockwell, Robert P., and Donka Minkova. 1988. The English vowel shift: problems of coherence and explanation. Luick revisited, ed. by Dieter Kastovsky and Gero Bauer. Gunter Narr Verlag: Tübingen.

Stockwell, Robert P., and Donka Minkova. 1997. On drifts and shifts. Studia Anglica Posnaniensia 31. 283–303.

Wolfe, Patricia. 1972. Linguistic change and the Great Vowel Shift in English, Berkeley: University of California Press.

DONKA MINKOVA

See also **English**

Greek, Modern

Modern Greek is the direct descendant of Koine Greek, the language of the New Testament, which in turn was based on the Attic dialect of Ancient Greek. It is thus part of the Greek or Hellenic branch of Indo-European. The language is called *eliniká* (Hellenic) by its modern speakers. Up until about the end of the last century, the spoken language was also referred to as *roméika* (Romaic) by the common people who spoke Greek as citizens of the Ottoman Empire before 1821 and then after the successful war of independence in 1821–1832 of the Kingdom of Greece. The modern Republic of Greece has Modern Greek as its sole official language. Modern Greek is also one of the two official languages of the Republic of Cyprus (Turkish is the other one). Greek speakers are located mostly in the Republic of Greece itself, with about 10,000,000 living there, but large numbers can also be found in Cyprus (c. 600,000) and parts of the diaspora (e.g. about 1,000,000 in Australia, chiefly in Melbourne and Sydney). Historically, Greek speakers have settled all over the eastern Mediterranean, in Southern Italy, along the Black Sea coasts, in Egypt, the Levant, Cyprus, and much of Asia Minor. This geographical spread continued throughout the Hellenistic period

and on through the Byzantine and Medieval periods, and is valid to some extent even into the Modern era, although most of the Greek inhabitants of Asia Minor (present-day Turkey) were removed to Greece (and many Greek-speaking Moslems from Greece to Turkey) after the population exchanges of the early 1920s in the aftermath of World War I. Within Greece, the greatest concentration of speakers, some 3,500,000 or more, lives in the greater Athens area alone, most of them speakers of the current standard language.

Faced with the difficult problem of distinguishing between dialects of a language as opposed to separate languages, the highly divergent modern form of Greek known as Tsakonian, still spoken in the eastern Peloponnesus (in Greece), could well be considered now as a separate language from the rest of Modern Greek, and the Pontic dialects once spoken along the Black Sea coast of Asia Minor but now spoken in many parts of Greece due to the 1923 population exchanges are divergent enough to warrant consideration now as a separate language from the rest of Greek. Similarly, modern Cypriot shows significant differences on all levels (phonological, morphological, and syntactic) that invite classification as a separate language, although this judgment is perhaps a more difficult one than in the case of Tsakonian or Pontic.

Still, it is customary to treat Modern Greek as a unified language with a range of dialects, much as was the case with Ancient Greek. While the dialect complexity of Ancient Greek was largely leveled out in Hellenistic times with the emergence of the relatively unified variety of Greek known as the Koine (see the chapter on Ancient Greek), the natural forces of language change led to new dialect diversity in the Byzantine period, with the modern regional dialects emerging after about the tenth to twelfth centuries AD. The main exception to this characterization is Tsakonian (as mentioned above), which derives more or less directly from the ancient Doric dialect, although with an admixture of standard Modern Greek in recent years; in addition, the Greek of Southern Italy, still spoken, for instance, in some villages in Apulia and Calabria, seems to have ancient Doric roots. The Pontic dialects (mentioned above) may derive more directly from the Hellenistic Koine.

Stemming from the later Byzantine form of the Koine, the major modern regional dialects are (following Newton 1972) Peloponnesian-Ionian, Northern, Cretan, Old Athenian, and Southeastern (including the Greek of the Dodecanese islands and, traditionally at least, Cypriot Greek as well). Peloponnesian-Ionian has formed the basis historically for what has become the contemporary Standard language, and is the basis for the Greek of modern Athens, as by far the leading population center in Greece; the Old Athenian dialect was the Greek of Athens before the 1821 War of Independence, and is still found elsewhere in Greece due to various resettlements.

A key aspect of the development of Modern Greek pertains to its external history, namely the fact that throughout post-Classical Greek, the language and its speakers were never able to escape the important cultural influence of the Classical Greek language and Classical Greece itself. The importance of Classical Greece—in the Mediterranean, the Balkans, parts of the Middle East, and even Western and Central Europe—meant that Classical Greek was taken as the prescriptive norm against which speakers of later stages of Greek generally measured themselves. This situation led to a 'two-track system' for the language, in which a high-style consciously archaizing variety that speakers and writers modeled on Classical Greek was set against a vernacular innovative variety. After the War of Independence from the Ottoman Empire in 1821 and the founding of a new nation-state of Greece, this distinction crystallized into a significant register and stylistic difference between what has come to be known as *Katharevusa* ('Puristic', literally '(the) purifying (language)') as the high-style variety associated with official functions, i.e. those pertaining to government, education, religion, and such, and *Dimotiki* ('Demotic', literally '(the) popular (language)') as the language of the people in ordinary, day-to-day, mundane affairs. This sociolinguistic state of affairs was the basis for the formulation of the notion of *diglossia* (Ferguson 1959), and struggles between advocates of each type of Greek, carrying with them certain social attitudes and political positions, continued throughout most of the twentieth century. After a number of governmental acts and actions in 1976, *Dimotiki* became the official language, and the diglossic situation was resolved, at least from an official standpoint. Significant for understanding variation in Greek is the fact that all throughout, both the official and unofficial periods of diglossia, speakers' usage was actually somewhat mixed, with borrowing between the two varieties, especially with Puristic forms incorporated into Demotic. The present state of Demotic, what has emerged as 'Standard Modern Greek,' hereafter SG, based on the everyday Greek of the largest city and capital of Greece (Athens), reflects a number of such borrowings from *Katharevusa*, involving both grammar (morphology and syntax) and pronunciation, as well as the lexicon, as discussed below.

Also relevant along with these stylistic/register differences is the effect of orthography. There is a long tradition of written forms of Greek, with the familiar Greek alphabet being the most enduring writing system for the language; as is so often the case, written forms tend toward the conservative, especially with respect to

the representation of pronunciation. There is thus within Greek, especially regarding phonology, a basis for influence from the written language, and the potential for variation therefrom. Moreover, spelling reforms of the late 1970s and early 1980s, leading to the so-called *monotoniko* ('monotonic') system, changed certain aspects of Greek orthography, in particular doing away with several phonetically irrelevant accent marks and diacritics that reflected Ancient Greek orthography; still, the old orthography can be encountered in books published before 1981 and in private use (e.g. personal letters), so that there is variation to be found in the form of written Greek even today.

What the long-term diglossia and associated influence from a written language have meant for Greek is the emergence of dialect differences that are not just regional (geographic) in nature. Rather, there are important socially based distinctions that have been fed by diglossia and by associations between conservative social and political attitudes and conservative linguistic usage on the one hand, and progressive attitudes and innovative linguistic usage on the other. Thus, within Greek one has to reckon with mixing of varieties and borrowing among them of both a regional and stylistic/social nature.

Other types of socially based variation can be found too, although beyond the omnipresent one based on the Katharevusa vs. Dimotiki distinction. From a functional standpoint, mention should be made of the existence of certain institutionalized trade jargons, e.g. that of coppersmiths, and several varieties of 'disguised languages' (e.g. one involving switching of syllables in a word with some distortions of vowels). Especially well known in this regard is καλιαρντά/kaliar^θ da/, the dialect of the gay community that is characterized especially by a large number of Turkish loanwords and divergent meanings for SG words (see Petropoulos 1971). One might also mention here conventionalized child-language forms (e.g. with s/z for SG dental fricatives T/Δ□, and various lexical items, as for bodily functions) that all (adult) speakers know and are able to use in appropriate situations (e.g. talking with young children).

Also of importance for the issue of the mixing of varieties in Greek is the presence of other languages in Greece and in the territory surrounding Greece in the Balkans down through the ages and even into modern times. These circumstances have led to the steady entry of numerous foreign words into Greek over the years, from Balkan, Middle Eastern, and more recently Western European languages, setting the stage for variation in the use and integration of loanwords on the part of Greek speakers. In the modern era, there have been periods of reaction against the influx of loanwords, with sometimes Italian but especially Turkish words being the prime targets for purging and replacement by 'native' Greek elements. These efforts have met with varying degrees of success, but in any case there are still large numbers of Turkish words in the language today, especially at the most colloquial and everyday levels of usage.

Thus, for a number of historical reasons having to do in large part with the geographic distribution of Greek speakers and with the particular circumstances of the relationship of later Greek speakers with their cultural past and heritage, Modern Greek today shows considerable variety in its realizations. Regional differences cut across social differences, and all this has come despite the existence in most periods of various strong centralized standard forms of the language (e.g. archaizing varieties in Medieval and early modern times, the demotic standard of today, etc.) that have provided norms for prescriptive usage.

As far as the linguistic structure of SG is concerned, it would perhaps be useful to draw a quick comparison between the most widely studied Classical Greek (see entry) and its evolution into Modern Greek. In terms of its phonology, Ancient Greek voiceless aspirated stops /t^h/ (θ), /p^h/ (φ), /k^h/ (χ) and voiced unaspirated stops /d/ (δ), /b/ (β), /g/ (γ) have changed to voiceless fricatives /T/ (T), ,/f/ (φ), /x/ (χ) and voiced fricatives /Δ/ (δ), /v/ (β), and /⊗/ (γ), respectively, making SG a language with a very rich fricative inventory. Vowel length is lost and all of the Ancient Greek diphthongs have been monophthongized in the modern language (e.g. AG /ei/, /oi/ becoming SG /i/, AG /ai/ becoming SG /e/, etc.). Furthermore, many AG vowels and diphthongs (e.g. /v/, /η/, /vι/, etc.) ended up pronounced as /i/ in SG, a phenomenon sometimes referred to as iotacism. SG has also lost pitch accent, which has been replaced by lexical (dynamic) stress. Aspiration /h/ has also been lost, and voiced stops have been developed anew, usually arising from voicing of voiceless stops when found after nasals or from borrowings (e.g. AG /pente/ 'five' giving SG /pe(n)de/ or /duvari/ 'wall' from Turkish). In terms of syntax, the modern language retains the free word-order capabilities of AG since most of the morphology and the inflections have been retained. Adjectives normally precede nouns, and definite articles have to accompany proper nouns like in Ancient Greek (e.g. *o giorgos* 'the George'). The dative case has been completely replaced by genitive and accusative and many AG monolectic verbal forms have been replaced by periphrastic constructions (e.g. AG *eureka* 'I have found' has become SG *exo vri*). The infinitive form of the verb has also been lost, reflecting perhaps a Balkan pattern of language contact.

References

Browning, Robert. 1983. Medieval and modern Greek. Cambridge: Cambridge University Press.

Holton, David, Peter Mackridge, and Irene Philippaki-Warburton. 1997. Greek. A comprehensive grammar of the modern language. London/New York: Routledge.

Horrocks, Geoffrey. 1997. Greek. A history of the language and its speakers. London/New York: Longman.

Joseph, Brian D., and Irene Philippaki-Warburton. 1987. Modern Greek. London: Croom Helm Publishers.

Mackridge, Peter. 1985. The modern Greek language. Oxford: Clarendon Press.

Mirambel, André. 1939. Précis de grammaire élementaire du grec moderne. Paris: Société d'éditions 'Les Belles Lettres'.

Newton, Brian. 1972. The generative interpretation of dialect. Cambridge Studies in Linguistics, Vol. 7. Cambridge: Cambridge University Press.

Petropoulos, Elias. 1971. Kaliarnta: Erasitechnike Glossologike Ereuna. Athens: Digamma.

GIORGOS TSERDANELIS

Greenberg, Joseph Harold

Joseph H. Greenberg was one of the most original and influential linguists of the twentieth century. Educated at a time when there were virtually no linguistics departments, he pursued studies in classics, anthropology, Indo-European comparative linguistics, and philosophy. Greenberg also shaped the empirical scientific study of linguistics, emphasizing the examination of a broad range of languages across the world in the development of theories of language structure, language change, and language classification.

Greenberg's earliest major work was on the genetic classification of the languages of Africa. Prior to Greenberg's work, the languages of Africa had been classified into five families, but the classification used inappropriate linguistic evidence and also nonlinguistic evidence. Greenberg established three fundamental principles in classifying languages. The first is to exclude typological linguistic features. Typological features are patterns of sound only—such as the presence or absence of tones—or of meaning only—such as the presence or absence of sex gender distinctions. The most reliable evidence for genetic classification is the pairing of sound and meaning, especially in grammatical inflections and basic vocabulary. The second principle is to exclude nonlinguistic evidence, such as skin color or cultural traits. The third principle is the simultaneous comparison of the vocabulary and inflections of a large number of languages (mass comparison, later called multilateral comparison).

Greenberg used these principles to classify the languages in Africa into four families: Afroasiatic, Nilo-Saharan, Niger-Kordofanian, and Khoisan. His classification was strongly criticized by established African scholars in Europe at first, but it was ultimately accepted by virtually all African linguists.

While working on the African classification through the 1950s, Greenberg also turned his attention to language classification in Oceania, the Americas, and Eurasia. He published only preliminary results at that time, but he continued to collect evidence, and assembling that evidence took up the last 20 years of his life. In 1971, he published an article presenting evidence that the languages of Papua New Guinea that do not belong to the Austronesian family—the so-called Papuan languages—and other non-Austronesian languages spoken as far afield as the Andaman Islands and Tasmania form a single large family, Indo-Pacific, with 14 major branches and a number of languages of uncertain affiliation. The Indo-Pacific hypothesis was, however, largely ignored by specialists in the Papuan languages. Although few linguists working in the area currently believe that all Papuan languages belong to a single family, a large grouping called the Trans-New Guinea Macrophylum has been proposed by some, and a number of its branches, as well as other accepted Papuan families, correspond to branches of Greenberg's Indo-Pacific family.

In 1987, Greenberg published a book presenting evidence that all of the indigenous languages of the Americas other than the Na-Dene family in the Pacific Northwest and the Eskimo-Aleut family in the Arctic belong to a single family, Amerind, with 11 major branches. This proposal has also been strongly criticized and stimulated a lengthy debate on the methods and principles of linguistic genetic classification, to which Greenberg contributed until his death.

The central criticisms of Greenberg's methods are that a genetic family cannot be established without reconstructing the ancestral language of the family; without doing so, it is impossible to determine whether resemblances in inflections and vocabulary are due to chance or borrowing, and whether the results are affected by errors in the transcription of the data. Also, since reconstruction is progressively more difficult as time depth increases, one cannot prove genetic hypotheses for families older than about 8,000 years. The conclusion of

413

most of Greenberg's critics, therefore, has been not so much that Greenberg's hypotheses are wrong, but that they are not provable in principle.

Greenberg argued in response that one cannot reconstruct a family without a prior hypothesis as to which languages form the family, since the number of possible groupings of even a small number of languages is astronomically high; and that genetic classification is only the first step, to be followed by reconstruction. Greenberg also argued that the probability of many languages simultaneously displaying similarities in form and meaning is so high that chance resemblances (and the interference of errors) decrease in likelihood. Greenberg also argued that borrowing displays specific probabilistic patterns that can be detected, and that differences in the relative stability of basic vocabulary and grammatical inflections mean that one can identify families older than 8,000 years.

The debate is not yet resolved, in part because it requires the use of quantitative techniques not yet applied to the problem. However, one can safely say that Greenberg's work has revived a subfield of linguistics that had been dormant for several decades, and has stimulated new research in language classification, language contact, and historical linguistics. Greenberg's work has also brought linguistics into contact with anthropology and genetics. Greenberg's three-family linguistic hypothesis for the Americas found parallels in research on human dentition and genetics, and geneticists such as Luigi Luca Cavalli-Sforza have compared Greenberg's deeper linguistic classifications to genetic evidence on prehistoric human migrations.

Greenberg published evidence for a family in Eurasia consisting of seven branches: Indo-European, Uralic-Yukaghir, Altaic, Japanese-Korean-Ainu, Gilyak (Nivkh), Chukotian, and Eskimo-Aleut. This work, is as controversial as his Amerind hypothesis.

Greenberg's other major line of research, also begun in the 1950s, is in the area of typology and universals of language. Typology began as the study of the range of different structural types of languages, according to their grammar (morphology and syntax) or their sound system (phonology). Languages vary considerably in structure. Greenberg linked typology to the study of universals of language through the discovery that structural variation across languages was limited, and those limits can be described in terms of implicational universals. An implicational universal is a universal of language formulated as an if–then statement, as in 'If a relative clause precedes the noun in a language, then a demonstrative adjective also precedes'. An implicational universal describes a relationship between two structural properties of a language; but it also allows for cross-linguistic variation in language type. For example, the aforementioned universal allows for language types in which the demonstrative precedes and the relative clause follows, and for both modifiers to precede or follow; it only disallows the type in which the relative clause precedes and the demonstrative follows.

Greenberg's first major publication on language universals proposed a series of universals of word order and morphological categories. The paper used a sample of 30 languages and the proposed universals were inferred from the distribution of languages in the sample among possible structural types. Among other things, the paper introduced the language types based on the relative order of subject (S), verb (V), and object (O), such as SVO, SOV, etc. This paper is one of the most cited papers in linguistics, and Greenberg's word-order types have been used widely in all theoretical approaches to linguistic analysis.

Greenberg's method of deriving universals became the fundamental method of the typological approach to grammar, and the typological approach was often compared to the generative approach of Noam Chomsky. The typological approach to grammar is characterized by: cross-linguistic comparison of a sample of genetically and geographically diverse languages; classification of the languages into types based on their structural properties; the inductive derivation of language universals by examining attested vs. unattested (or very rare) language types in the sample; and explanation of the universals, usually by appeal to functional motivations (semantic, pragmatic, discourse-functional, or processing principles).

Greenberg's paper on word order was only one of many typological studies that he produced. Many of these studies pertained to sound structure: on the consonant–vowel dichotomy, consonant clusters, glottalized consonants, and word prosodic systems. Others were concerned with grammar: besides word order, Greenberg produced studies of numeral systems and of a typological interpretation of the concept of markedness and markedness hierarchies in inflectional categories.

In the 1960s, Greenberg became more interested in universals of language change, which he christened diachronic typology. Greenberg produced a series of seminal methodological papers in diachronic typology, and also a number of studies of universals of language change, including voiceless vowels, numeral classifier constructions, and gender marking. The last study helped begin the revival of the study of grammaticalization—the evolution of grammatical categories and constructions from independent words and syntactic patterns. Grammaticalization theory is now a central area of research in historical linguistics and typology.

Greenberg's seminal methodological and empirical papers in typology and universals, including diachronic typology, created the foundations of several major contemporary strands of linguistic research. Greenberg also made significant contributions to African linguistics,

sociolinguistics, and psycholinguistics. Although the fate of his genetic classifications is still being debated, Greenberg's legacy in the empirical, cross-linguistic study of language universals and language change is undoubtedly secure.

Biography

Joseph Harold Greenberg was born in Brooklyn, New York on May 28, 1915; he married Selma Berkowitz on November 23, 1940. He received his B.A. (1936) from Columbia University and Ph.D. (1940) for his dissertation on the influence of Islam on the Hausa, tutored by Melville Herskovits, Northwestern University. He was with the US Army Signal Intelligence Corps, 1940–1945; Assistant Professor, University of Minnesota—Minneapolis, 1946–1948; Assistant Professor, 1948–1953, Associate Professor 1953–1957, Professor, 1957–1962, Columbia University; Professor, Stanford University, 1962–1985; and Ray Lyman Wilbur Professor of social sciences in anthropology, 1971. He was also Director of African Languages and Area Center, 1967–1978 and co-editor of *Word*, 1950–1954. He was a Fellow of the National Academy of Sciences, the American Philosophical Society, and the American Academy of Arts and Sciences, and a Guggenheim Fellow in 1954–1955, 1958–1959, and 1982–1983. He was also Fellow, Center for Advanced Study in the Behavioral Sciences in 1959–1960 and 1965–1966, and Stanford Humanities Fellow in 1982–1983. He was the First Distinguished Lecturer, American Anthropological Association in 1970 and Hermann Collitz Professor, Summer Institute of the Linguistic Society of America 1987. He was President of the African Studies Association, 1964–1965; Chairman of the West African Linguistics Association, 1965–1966; and President of the Linguistic Society of America, 1977. He received the Haile Selassie Award for African Research in 1967, the New York Academy of Sciences Behavioral Sciences Award in 1980, and the National Academy of Sciences/Academy of Arts & Sciences Talcott Parsons Prize for Social Science in 1997. Greenberg died in Stanford, California on May 7, 2001.

References

Comrie, Bernard. 1989. Language universals and linguistic typology, 2nd edition. Chicago, IL: University of Chicago Press.

Croft, William. 2001. Joseph Harold Greenberg (1915–2001). Language 77. 815–30.

————. 2003. Typology and universals, 2nd edition. Cambridge: Cambridge University Press.

Croft, William, Keith Denning, and Suzanne Kemmer (eds.) 1990. Studies in typology and diachrony for Joseph H. Greenberg. Amsterdam and Philadelphia, PA: John Benjamins.

Greenberg, Joseph Harold. 1957. Essays in linguistics. Chicago: University of Chicago Press.

————. 1963. The languages of Africa. Bloomington: Indiana University Press.

———— (ed.) 1963. Universals of grammar. Massachusetts: MIT Press Cambridge.

————. 1966. Language universals, with special reference to feature hierarchies. The Hague: Mouton.

————. 1968. Anthropological linguistics. New York: Random House.

————. 1971. Language, culture and communication: essays by Joseph Greenberg. Stanford: Stanford University Press.

————. 1974. Language typology: a historical and analytic overview. The Hague: Mouton.

————(ed.). 1978. Universals of human language, 4 vols. Stanford: Stanford University Press.

————. 1987. Language in the Americas. Stanford: Stanford University Press.

————. 1990. On language: selected writings of Joseph H. Greenberg. Stanford: Stanford University Press.

————. 2000/2002. Indo-European and its closest relatives: the Eurasiatic language family, 2 vols. Stanford: Stanford University Press.

————. 2005. Genetic linguistics: theory and method. Oxford: Oxford University Press.

Heine, Bernd, and Derek Nurse (eds.) 2000. African languages: an introduction. Cambridge: Cambridge University Press.

Newman, Paul. 1995. On being right: Greenberg's African linguistic classification and the methodological principles which underlie it, Bloomington, IN: Institute for the Study of Nigerian Languages and Cultures, African Studies Program, Indiana University.

WILLIAM CROFT

See also **Afroasiatic; Grammaticalization; Typology**

Grice, H. Paul

In the early period of his career, H. Paul Grice participated in Oxford's ordinary language philosophy school. Some of his work is critical of particular analyses of ordinary language philosophers, such as Strawson's analysis of logical connectors (e.g. 'and', 'or', and 'if') and Austin's 'no modification without

aberration' thesis. Nonetheless, Grice retained a distinct interest in 'conceptual analysis' (i.e. the analysis of the possible uses of a linguistic expression) as a privileged, albeit not exclusive, method of philosophical analysis.

The most significant contributions of Grice to linguistics revolve around a broad attempt to reduce semantics to intentionality, via the notions of *meaning*$_{nn}$ and *implicature*. Meaning$_{nn}$ (nonnatural) is opposed to natural meaning. While natural meaning is independent of any intentionality (e.g. the presence of smoke means that there is a fire), meaning$_{nn}$ is defined specifically in terms of the speaker's intention that the hearer recognize his or her intention to mean something. The original formulation of the idea underwent several reformulations under the pressure of several counterexamples (Schiffer 1972), but the following is particularly perspicuous: the utterer (U) of an utterance *x*, meant something, if U uttered *x* intending, given an audience (A),

(1) A to produce a particular response *r*
(2) A to think (recognize) that U intends (1)
(3) A to fulfill (1) on the basis of his fulfillment of (2) (Grice 1989:92).

This M-intention (Grice 1989:105) or 'reflexive intention' (Searle 1969:47) is taken to be the defining feature of meaning$_{nn}$. A significant issue, which seems to have received very little attention (Neale 1992), is that all reformulations of the M-intention fall victim to counterexamples in which the speaker's motives are deceptive. Grice suggests (1989:302–3) blocking out 'sneaky' intentions by stipulating that one cannot mean deceptively. This provides an interesting link between the concept of meaning$_{nn}$ and the cooperative principle, specifically the maxim of Quality (see below).

The conventional nature of the speaker's reliance on meaning$_{nn}$ has been pointed out (Lewis 1969). A potentially problematic infinite regression of 'mutual knowledge' (Lewis 1969, Schiffer 1972) has also been noted. The issue revolves around the fact that the hearer knows that the speaker intended a proposition P, but the speaker must know that the hearer knows that the speaker intended P, and so on. Many have challenged the original definition of the nature of the M-intention on the basis of this alleged infinite recursion. Various solutions have been proposed to this problem. Others have challenged the exclusive reliance on intentions, without use of the conventional (literal) meaning of the utterance (Searle 1969). Grice, in fact, argues that conventional linguistic meaning can be reduced to intended speaker meaning. The meaning$_{nn}$ thesis remains controversial and has in fact been repudiated by some of its early proponents (Schiffer 1987). Others have used Grice's meaning$_{nn}$ and Lewis's con-

ventionality as the basis for a general theory of language in a behaviorist mold (Bennett 1976) or as the foundation of a cognitive theory (Sperber and Wilson 1986).

Grice's intentionalism is not limited to meaning$_{nn}$; on the contrary, he distinguishes between what is said in an utterance (roughly, literal meaning) and what is implied by it. For the latter notion, Grice introduced the neologism *implicature*. Recent work has questioned the division said/implied and introduced a third term (variously labeled *explicature/impliciture*) to denote the aspects of the meaning of an utterance that while not explicitly stated, and hence derived pragmatically, are nonetheless not part of implicature.

In its simplest form, the notion of implicature can be defined as some proposition that while not explicitly stated by the speaker is nonetheless intended by him or her and is understood by the hearer as such. Implicatures may be context-free (generalized) or context-sensitive (particularized). In the former case, they apply in any utterance of a sentence, regardless of the circumstances/context. In the latter, they are sensitive to the context. Both types of implicatures are governed by the cooperative principle (CP), arguably Grice's best-known contribution to linguistics. The CP consists of a general statement:

> make your contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged

and of four maxims:

- The maxim of Quality

    try to make your contribution one that is true; specifically
    do not say what you believe to be false
    do not say that for which you lack adequate evidence

- The maxim of Quantity

    make your contribution as informative as is required for the current purposes of the exchange
    do not make your contribution more informative than is required

- The maxim of Relevance

    make your contribution relevant

- The maxim of Manner

    be perspicuous and specifically
    avoid obscurity
    avoid ambiguity
    be brief
    be orderly (Grice 1989:26–27).

The exact relationship between the general principle and the maxims has been the object of controversy. The generally accepted opinion is that the maxims instantiate the principle. Implicatures can be drawn on the basis of the CP by either following its rules or by deliberately and conspicuously not doing so (called *flouting*). If the speaker follows the maxims, then the hearer can assume that he or she is being truthful, relevant, etc. For example, the utterance of 'Mary won the Nobel prize' implicates that the speaker has adequate evidence that she did so. On the contrary, if the speaker is obviously not following the maxims, and makes no attempt at concealing this, then the hearer can seek an explanation. For example, as the phone rings, a speaker says 'I am not here'. Since the violation of the CP is obvious, the hearer considers the utterance relevant to the context (the incoming phone call) and therefore assumes that the speaker does not wish to talk to whoever is calling.

Reactions to Grice's CP have ranged widely. There have been total rejections based on accusations of naiveté or on the grounds that, simply put, people do not generally behave as Grice's CP would predict. These objections can be countered by the observation that Grice does not claim that people are always cooperative. Others have claimed that the CP is prescriptive (i.e. tells speakers how they *should* behave). This view is in error, as stated in the literature, but deserves reconsideration in light of Grice's views on morality. It has been claimed that the CP has been empirically falsified by anthropological evidence and that therefore its universal nature has been falsified, but these claims have been repeatedly refuted. More moderate, but still negative, reactions have embraced the general impetus of the theory but claimed that the CP and/or the maxims were too specific and/or too general and/or too vague. A related criticism is that no explicit procedure is provided to calculate the implicatures. Several authors have remarked that the CP in its original wording is limited to declarative statements and have offered rewordings that broaden its scope to nonindicative and nonassertive utterances. Considerable effort has gone into attempts at formalizing the individual maxims, especially the maxim of Relevance, within the framework of Relevance Theory, and the maxim of Quantity. The judgment on the effectiveness of these attempts is also mixed. Many have tried to determine the criteria for deriving conversational implicatures and distinguishing them from other semantically related propositions, such as presuppositions. It has been suggested to reduce the maxims to two or even one principle, often based on processing constraints, thus linking to the functionalist tradition (e.g. the principle of least effort). Conversely, there have been many proposals to augment the CP with more maxims, or even to proliferate principles. Recently, a proposal to augment the CP with a set of heuristics, based on the maxims, has been presented (Levinson 2000). These heuristics generate *generalized conversational implicatures,* which are default implicatures that hold in most contexts.

Overall, however, the critical voices have been drowned out by the sheer mass of those who have silently accepted and incorporated a more or less faithful Gricean distinction between what is said and what is implicated by an utterance. It is fair to say that the distinction has entered the realm of the basic linguistic notion that is part of the curriculum taught to beginners.

In the Californian phase of his career, Grice focused his attention toward ethics (1991). This aspect of his work is not unrelated to the research on the problems of meaning, although most linguistic applications have neglected to assess the connection between the two aspects of Grice's thought. For example, Grice attempted to provide a nonrelativistic, metaphysical foundation for value judgments, such as those of morality. This led Grice to the adoption of a substantialist view of rationality. A substantialist (i.e. noninstrumental) rationality is concerned not only with the rationality of the means to an end but also with the rationality and morality of the ends themselves. The connection with the cooperative principle is thus clear. Another example of connection between Grice's ethics and his semantics is his insistence that the reflexive intention of meaning$_{nn}$ is rational and hence evaluative (1989:283–303).

Biography

Herbert Paul Grice (1913–1988) began his career in his native England where he graduated from Oxford's Corpus Christi College in 1935 and received an M.A. in philosophy from Merton College in 1938. He became a fellow of St. John's College, where he taught from 1939 until 1967 when he moved to Berkeley, California. In 1967, Grice delivered the William James lectures (published partially in 1975 and fully published in 1989), which were widely circulated in manuscript. The widespread interest in the idea of implicature dates back to these lectures. He taught at Berkeley until his retirement in 1979. Between 1980 and 1983, he taught at the University of Washington. In 1983, Grice delivered the Carus lectures on the conception of value, published posthumously as Grice (1991). Significant parts of Grice's writing remain unpublished. The move from Europe to the United States parallels a shift of interests in his work. In the early part of his career, Grice adhered to and contributed to the development of the ordinary language philosophy of Austin, Ryle, and other Oxford

philosophers, although he always maintained a distance from this school. During this period, he worked primarily on problems of meaning. In the second phase of his work, he moved toward issues of ethics and metaphysics. Grice's influence on linguistics is largely restricted to the first phase of his work.

References

Bennett, Jonathan. 1976. Linguistic Behaviour. Cambridge: Cambridge University Press.

Cosenza, Giovanna. 1997. Intenzioni, significato, comunicazione. La filosofia del linguaggio di Paul Grice. Bologna, Italy: Clueb.

Grandy, Richard E., and Richard Warner. 1986. Philosophical grounds of rationality: intentions, categories, ends. Oxford: Clarendon.

Grice, Paul. 1989. Studies in the way of words. Cambridge, MA: Harvard University Press.

_____. 1991. The conception of value. Oxford: Clarendon.

Hall, Kira, et al. 1990. Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society. February 16–19, 1990. General Session and Parasession on the Legacy of Grice. Berkeley, CA: Berkeley Linguistics Society.

Levinson, Stephen C. 1983. Conversational implicature, Ch. 3. Pragmatics, 97–167. Cambridge: Cambridge University Press.

Levinson, Stephen C. 2000. Presumptive meanings: the theory of generalized conversational implicature. Cambridge, MA: MIT Press.

Lewis, David K. 1969. Convention: a philosophical study. Cambridge, MA: Harvard University Press.

Neale, Stephen. 1992. Paul Grice and the philosophy of language. Linguistics and Philosophy 15. 509–59.

Schiffer, Stephen R. 1972. Meaning. Oxford: Clarendon.

Schiffer, Stephen R. 1987. Remnants of meaning. Cambridge, MA: MIT University Press.

Searle, John. 1969. Speech acts: an essay in the philosophy of language. Cambridge: Cambridge University Press.

Sperber, Dan, and Deirdre Wilson. 1986. Relevance: communication and cognition. Oxford: Blackwell, 2nd edition, 1995.

SALVATORE ATTARDO

See also **Meaning**

Grimm, Jacob

Together with his brother Wilhelm, Jacob Grimm is perhaps the most influential scholar in the field of Germanic linguistics in the nineteenth century. The sons of a lawyer, both brothers studied law in Marburg, and became interested in the history and traditions of Medieval Germany. Jacob Grimm's first publications, all of which issued from close collaboration with Wilhelm (a constant companion to his brother's life and work), were editions of Medieval texts, and the famous collection of fairy tales (*Kinder-und Hausmärchen*, published starting with 1812).

Research on ancient texts and on oral tradition finally led Jacob Grimm to develop an interest in language change, which at his time mostly meant tracing the history of particular words. Very soon, Jacob Grimm turned to grammatical description. In 1819, he published the first volume of his most important work, the *Deutsche Grammatik*. The title should not be taken as meaning 'German grammar' in a strict sense: the word *deutsch* here is used by Grimm as including all Germanic languages, at all their documented stages.

Grimm conceived of language as an ever-changing phenomenon, which can be studied only through empirical observation. Consequently, he did not only include all early stages of the Germanic languages in his grammar, but also referred to correspondences in other ancient Indo-European languages. For this reason, his work acquired primary importance for Indo-European linguistics, and he is nowadays regarded as one of the founders of Indo-European comparative studies.

Although the *Deutsche Grammatik* included four volumes (after the first, the others followed in 1826, 1831, and 1837), what earned Jacob Grimm his fame in the field of Indo-European linguistics is constituted by the first chapter of the first volume, a later addition to the 1822 second edition, devoted to sound structures. In this chapter, called in the German original text *Von den Buchstaben*, Jacob Grimm set out to provide a description of correspondences in the pronunciation of words, based on comparative research and following his conviction that pronunciation is autonomous from the rest of grammar. Grimm decided to add this important chapter after reading Rask's Icelandic grammar, in which Rask described what was later to be known as the 'first sound shift' (*erste Lautverschiebung*) or 'Grimm's Law'.

Although Rask must be credited with having been the first to understand the sound change that now bears Grimm's name, Grimm had already demonstrated his own insight into sound change in the description of particular vowel alternations (*Umlaut* and *Ablaut*) he

gave to G.F. Benecke, the editor of many Medieval texts, in 1816. Grimm observed that, much in the same way as certain Gothic consonants had undergone a one-step change away from the related Greek and Sanskrit, certain High German consonants had moved one step more in the same direction, as exemplified in the following table:

Greek became	p	b	ph	t		d	th	k	g	kh
Gothic which became	f	p	b	th	t	d		h	k	g
High German	b/v	f	p	d	z	t		g	ch	k

Hence, Grimm described consonant shift as a recurrent tendency of Germanic, and called the sound shift from Gothic to High German *zweite Lautverschiebung,* or 'second consonant shift'.

Research on the history of Germanic and other Indo-European languages did not exhaust Grimm's interest in language. Apart from further writings in the field (among which it is worth remembering the *Deutsches Wörterbuch*, started in 1854, again with the collaboration of Wilhelm, and never finished), Jacob Grimm also published an essay on the origin of language (*Über den Ursprung der Sprache*) in 1851. As habitual in the first part of the nineteenth century, Grimm conceived of language evolution as a process of decay from the 'perfect' ancient Indo-European languages with complex systems of inflection to the 'degraded' modern languages that use simpler structures with independent words instead. However, his deep empirical knowledge of language led him to understand that language change did not necessarily impoverish languages, but that, on the contrary, the loss of old categories is compensated by the creation of new ones, and that the latter are not necessarily 'worse' than the former, even if their implementation does not involve the same type of complexity. Thus, Grimm, in a very original way, and partly contradicting his own initial statements, ends up praising English as the perfect combination of Germanic and Latin elements, which make it perhaps the most powerful human language.

Jacob Grimm did not lead the life of a retired scholar: he was active in the political events of his age, and in 1848 he was, for a short time, member of the parliament of the German Republic. Especially because of his research on the history and traditions of the Germanic peoples, he was considered, already in his times, one of the fathers of German culture.

Biography

Jacob Grimm was born in Hessen (Germany) on January 4, 1785. He pursued juridical studies in Marburg under Friedrich von Savigny (1779–1861), then assistant of Savigny in Paris in 1805. After various appointments in Kassel, he returned to Paris as a representative of the Prussian diplomacy and took part in the Vienna Congress in 1815. From 1816 to 1829, he worked in Kassel as a librarian. From 1830 to 1837, his first appointment was as a professor in Göttingen. In 1837, he lost his position as a consequence of having protested against the revocation of the constitution and returned to Kassel. In 1840, he became a member of the Prussian Academy of Sciences in Berlin. He moved to Berlin in 1841. He died in Berlin on September 20, 1863.

References

Collinge, N. E. 1985. The laws of Indo-European. Amsterdam and Philadelphia, PA Benjamins.

Grimm, Jacob. 1822–1837. Deutsche Grammatik I-IV, 2nd edition. Göttingen; reprinted Hildesheim: Olms, 1967.

————. 1970. Geschichte der deutschen Sprache, 2 vols. Leipzig: Weidmann, reprinted Hildesheim: Olms.

————.1864–1890. Kleinere Schriften, 8 vols. Berlin: Gütersloh, Dümmler, Bertelsmann.

Morpurgo Davies, Anna. 1996. Nineteenth century linguistics. vol. 4, London: Longman. History of linguistics, ed. by Giulio Lepschy.

Sebeok, Thomas A (ed.) 1966. Portraits of linguists. Bloomington and London: Indiana University Press.

SILVIA LURAGHI

See also **German; Gothic**

Guaymí and Chibchan Languages

The Chibchan languages are spoken in a wide area extending from Northeastern Honduras, through the Atlantic Coast of Nicaragua, most of Costa Rica, Panama, Colombia, and the West of Venezuela. This language family constitutes the largest family in Central America and Colombia, and the best-studied family of

the linguistic area known as Intermediate. The name *Chibcha* stems from the civilization established around Bogotá (Colombia) at the time of the Spanish conquest; its language, also known as Muisca, was a member of this family. The term Chibcha has been used in many cases as a 'ragbag' for languages whose affiliation to other (larger) families is uncertain. In the past, languages as disparate as Paez (Ecuador), Tarasco (Mexico), members of other families (e.g. Carib, Aztec), and even languages from as far as Chile (Atacama) and Argentina (Allentiac) have been labeled 'Chibcha'. Recent, modern glottochronological studies, in which parallelisms between languages are used to date a potential ancestor language, have yielded a less extensive family.

The geographic distribution of the family has given rise to some controversy. The fact that only a few members are spoken in the peripheral north has given rise to two views: the Theory of North Migration and the Centrifugal Expansion Theory. The former posits Colombia as the epicenter of the Chibchan languages, from where northbound migratory movements took place in relatively late times (the name *Chibcha* applied to the whole family mirrors in this view). Glottochronological studies, however, do not seem to back up this position (which is not based on linguistic data anyway), but tend to favor the latter, according to which the original Chibchan territory was the present-day border zone between Costa Rica and Panama, at the Talamanca mountain range, which cross-cuts that border; from there, migratory movements both northbound and southbound took place. Recent anthropological research tends to back the Centrifugal Expansion Theory. The Chibchan languages are thus Central American languages.

There are 24 members in the family, eight of which are now extinct: Huetar (Costa Rica); Dorasque and Chánguena (Panama); Muisca, Duit, Atanques, Catío, and Tairona (Colombia). Others are facing imminent extinction, e.g. Paya (Honduras), Rama (Nicaragua), Guatuso, and Boruca (Costa Rica), having less than 300 speakers. The majority of the living languages have an average of 3,000 speakers and are, at least temporarily, not seriously threatened by extinction.

The living members, their geographic distribution, and the number of speakers follow (with alternative names in parentheses):

Paya (Pech)	Honduras	600
Rama	Nicaragua	24
Guatuso (Maleku)	Costa Rica	365
Bribri	Costa Rica	6,000
Cabécar	Costa Rica	2,500
Boruca (Brunca)	Costa Rica	4
Teribe (Naso)	Costa Rica, Panama	1,500
Guaymí (Ngäbére,Movere)	Costa Rica, Panama	150,000

Bocotá	Panama	2,500
Cuna	Panama, Colombia	70,000
Ika (Bíntucua, Arhuaco)	Colombia	7,000
Damana (Arsario, Malayo, Guamaca)	Colombia	1,500
Kogi (Cágaba)	Colombia	8,000
Tunebo (UwCuwa)	Colombia	1,800
Chimila	Colombia	900
Barí (Motilón)	Colombia-Venezuela	2,000

There seem to be light differences between the languages of Colombia and those of Central America in a few respects. Languages in Colombia seem to use more auxiliaries and agreement markers. Clause connectors as part of verbal morphology are absent in Central America (except in Rama), but present in Colombia. All these differences point to a more elaborate verb structure in Colombia than in Central America.

Central America is a convergence zone. Two cultural and linguistic areas meet there, that known as Mesoamerica and the so-called Intermediate Area. In general, the Mesoamerican languages (Mayan, Aztec) exhibit a verb-initial basic word order pattern of V(erb)–S(ubject)–O(bject) or VOS, they have prepositions, a possessed–possessor and an adjective–noun order; the languages of the Intermediate Area (including Chibchan), on the other hand, are SOV, they have postpositions, and possessor–possessed and noun–adjective orders.

Structural Features

Basic Word Order

The default word order in Chibchan is exceptionless: Subject–Object–Verb. The other possible orders occur under the following conditions: (a) use of pronouns, (b) subject occurring in postverbal position; and (c) implied subject. Option (a) provides the widest range of possibilities. If the verb carries agreement markers identifying both subject and object, the word order is relatively flexible: s–o–V (see (1) below), o–s–V (see (2) below), or o–V–s (see (3) below), but not s–V–o:

(1) *mu-nu-pash-ka*   DAMANA
  you-me-hit-FACTUAL
  'You hit me'

(2) *ma ni-sung-u*   RAMA
  you I-see-PAST
  'I saw you'

(3) *mi-tšua-na-rua*   IKA
  you-see-far-I
  'I saw you'

As for alternative subject position (b), in some languages it can be placed postverbally; the object

Geographic Location of the Chibchan Languages.

appears in sentence-initial position for highlighting purposes, as in Cabécar:

(4) *tkabë su-wá*   *ét-aba*   *yís të*
     snake
     see-PERFECTIVE
     one-long   I
     'A snake I saw'

The implied subject realization (c) is illustrated in (7) below.

Word Classes and Parts of Speech

There are two basic types of lexical words in Chibchan, around which most of the grammatical features and categories revolve; these are nouns and verbs. A second set of word classes is formed by adjectives and adverbs; these are second to the former in that although there are clearly identifiable adjectival (colors and basic qualities) and adverbial (time and manner) roots, their number is not that large, and most other adjectival and adverbial concepts are expressed by other means (deriving adjectives/adverbs from noun or verb roots, using phrasal expressions, etc.).

Nouns and Nominal Morphology

The most common type of determiners are numeral classifiers, forms used to categorize nouns exclusively in the context of quantification; these are found in the languages of Costa Rica and Panama. The existence of numeral classifiers in the Chibchan languages is the result of the evolution of certain classes of nouns used to form lexical compound nouns in the assumed common ancestor (proto) language; the position of the numeral form in relation to the noun (prenominal vs. postnominal) in Proto-Chibchan determined the for-

mal status of the class markers either as prefixes, as in Teribe, or as suffixes, as in Bribri, Kogi, or Cabécar. Most languages with noun classifiers have around five classes—some have more: animate, long, round, thin, human; in most languages, those classifiers are free forms, i.e. words that can stand independently. Teribe, for example, has six classes: (a) prototypically animate objects; (b) round objects; (c) wide objects; (d) long objects; (e) long and wide objects; and (f) objects that can be counted in plots. The expression of each of the classes is effected through prefixes attached to a numerical base. For example, the base for the number one is *ara*, to which the class marker prefix is added: *kl-ara* (animate), *kw-ara* (round), *k-ara* (wide), *pl-ara* (long), *kwan-na* (long-wide), and *kri-na* (plots).

The Chibchan languages use demonstrative determiners, which are free forms, to express spatial relationships. The languages can be grouped into those that have three degrees of spatial distance (Teribe, Rama, Cuna, Kogi, Damana) and those that have two degrees (Boruca, Ika, Tunebo). These languages can also be divided in terms of the determiner's position relative to the noun, as prenominal (Rama, Guatuso, Boruca, Cuna) and postnominal (Teribe, Bocotá, Bribri, Cabécar); one language, Guaymí, allows both orders.

The Chibchan personal pronouns exhibit two basic traits; (a) with the exception of Rama, the free pronoun sets are formally indistinguishable for the core cases (subject and object), and (b) free personal pronouns are generally used for emphasis. This rather limited use of personal pronouns has to do with two other pervasive features of these languages, namely the use of implied subjects/objects and, in most members, cross-referencing mechanisms.

Where the word forms do not identify subjects vs. objects, verbal affixes help to disambiguate, as in (5), where the verbal suffix marks the subject:

(5)   *lanyo kwe*   TERIBE
      *midë-**rwa** llëme*
      story DEMONSTRATIVE know-we not
      'We don't know that story'

Another strategy is to use particles identifying the subject or topic, in combination with word order:

(6a) *At **ki** ba ishd-ra*   BORUCA
      I SUBJECT YOU see-PRESENT
      'I see you'
(6b) *Ba **ki** at ishd-ra*
      you SUBJECT I see-PRESENT
      'You see me'
(6c) *At ishd-ra ba-ng*
      I see-PRESENT you-TOPIC
      'You see me'

A very conspicuous feature of Chibchan is the existence of forms expressing the status of participants as topic (what the sentence is about) or focus (new information). Such markers combine with word order to create a wealth of foregrounding structures similar in function to such mechanisms as passives or left/right dislocations in other languages. In Teribe, the topic marker is li (7); there are two focus markers, *om*, *omgo* (8a, b), and one contrast marker ra (8a):

(7)   *tlɛ) lok ga 'walë kuzong ga walë **li***'
      say PLURAL that
      'woman then that woman TOPIC
      yo-y-dë'
      appoint-we-will'
      'Then they said, 'well then, if
      it's a woman, a [this] woman
      we will appoint'
(8a) *Kone kone **om** wuë, pero kone*
      *kone **om** wuë un llëme...*
      Some FOCUS eat, but some FOCUS
      eat all not ...
      *tawa **ra om** wuë llëbo un.*
      we CONTRAST FOCUS eat thing all.
      'Some eat THEM, but some
      [others] do not eat THEM...
      WE DO eat THEM ALL'
(8b) *E tle) ga e **omgo** twa-ydë*
      DEMONSTRATIVE say CONNECTIVE
      DEMONSTRATIVE FOCUS come-will
      'He said HE is the one
      who is coming'

Verbs and Verbal Morphology

T(ENSE)A(SPECT)M(OOD) systems are abundant and in some cases highly elaborated. Some languages do not overtly distinguish tenses; tense-marking languages are Rama, Guatuso, Cuna, Kogi, and Tunebo. Some tense-marking languages also have aspectual distinctions (Rama, Kogi); others (Tunebo) have only tense.

In most Chibchan languages, the most basic TAM distinctions are expressed by bound morphology (affixes), not by free-standing auxiliaries. Auxiliaries are mainly found in the languages of Colombia, as in Kogi:

(9)   *Nac gu-ngu-cu-aini-ki,*   *yatai*
                                        na-gu-cu-a
      come do-PAST-I-and-then,   sit CERTAINTY
                                        do-I-PAST
      'I came and then I sat down'

In some Central American languages, there are positional verbs, a closed set of linking verbs specifying location and manner of the location (standing, lying), as in Bribri:

(10) *ie? dur âula a*
      he POSITION.STAND classroom in
      'He's standing in the classroom'

Teribe has seven positionals (*löng*, *shäng*, *pang*, *buk*, *sök*, *jong*, *lok*). The remarkable fact about these positionals is that sentences expressing actions and movement are often ungrammatical if they lack such a positional:

(11a) *Tawa shro-no löng na*
      we arrive-PERFECTIVE
      POSITION.BE here
      'We came here'   (ungrammatical
                                        without *löng*)
(11b) *ëng wle) -no löng e wlo*
      each other meet-PERFECTIVE
      POSITION.BE DEMONSTRATIVE
      in order to
      'for us to meet'   (ungrammatical
                                        without *löng*)

Person Markers (Agreement)

The Chibchan languages fall into three groups in terms of agreement: (a) those that have no agreement (Bribri, Cabécar, Boruca, Bocotá, Guaymí, Cuna, Tunebo); (b) those with subject agreement (Rama, Teribe); and (c) languages with both subject and object agreement (Guatuso, Ika, Kogi). A special case is Damana, which has verbal agreement markers for the indirect (dative) object in addition to agreement for the subject and the direct object.

Grammatical Relations

Seven languages have nominative–accusative case-marking, where nominative case marks the subject and accusative the (direct) object: Paya, Rama, Boruca, Teribe, Cuna, Chimila, and Barí.

Ergative Systems

Eight languages (Guatuso, Bribri, Cabécar, Guaymí, Ika, Kogi, Damana, and Tunebo) use ergative–absolutive case-marking, where the subject of an intransitive verb patterns with the direct object of a transitive one (abso-

lutive), while the subject of a transitive verb is marked with ergative case.

(12a)   *ie$^{?}$-r*   *dìwö s$\underline{a}$-w$\underline{e}$*   BRIBRI
         she-ERGATIVE   sun(ABSOLUTIVE)   see-PERFECTIVE

         'She saw the sun'

(12b)   *dìwö*   *mìchò*
         sun(ABSOLUTIVE)   go
         'The sun goes'

An important property of some of these languages is that the ergative marker may only be used when it is needed for emphasis.

Active/Nonactive Systems

One language, Bocotá, appears to follow an active/nonactive pattern. Direct marking is used to mark the active subject by *no*:

(13a)   *Cha no*   *gliá*   *gúg-le*
         I   ACTIVE   leaves   burn-PERFECTIVE
         'I burned leaves'

(13b)   *Cha*   *gugé-du*

         I   burn-PERFECTIVE
         'I burned'

(13c)   *Gliá*   *gugé-du*
         leaves   burn-PERFECTIVE
         'Leaves burned'

References

Constenla, Adolfo. 1991. Las lenguas del Área Intermedia. San José: Editorial Universidad de Costa Rica.

Constenla, Adolfo. 1995. Sobre el estudio diacrónico de las lenguas chibchenses y su contribución al conocimiento del pasado de sus hablantes. Boletín del Museo del Oro (Colombia) 38/39. 13–35.

Quesada, J. Diego. 1999a. CHIBCHAN: with special reference to participant-highlighting. Linguistic Typology 3(2). 59–108.

Quesada, J. Diego. 1999b. Ergativity in Chibchan. Sprachtypologie und Universalienforschung 52(1). 22–51.

Quesada, J. Diego. 2000a. A grammar of Teribe (Chibchan, Panama). Munich: LINCOM-EUROPA.

Quesada, J. Diego. 2000b. Word order, participant-encoding and the alleged ergativity in Teribe: International Journal of American Linguistics (IJAL) 66(1). 98–124.

J. DIEGO QUESADA

Gullah

Gullah, sometimes also identified as *Geechee* or *Sea Island Creole*, is an American English creole spoken in coastal South Carolina and Georgia. It is not the only variety of American English that developed from contact with another language, but it is the only one that has been singled out as a creole. It shares several structural features with Caribbean English vernaculars that have also been called creoles. Like the latter, it developed on plantations where African slaves outnumbered the European indentured servants and other colonists with whom they interacted.

Gullah differs from African American Vernacular English (AAVE, 'Black English') in the socioeconomics of its development as well as in some of its structural features (including a speech melody that is closer to that of Bahamian and Caribbean Englishes). For instance, in Gullah one can form the progressive with the preverbal marker [də] (spelled *duh* or *da* in 'eye dialect'), as in *he duh/da talk(in)* 'he is talking', but not in AAVE. Gullah developed on the coastal rice fields, where the African slaves were in the majority from the early eighteenth century to the mid-nineteenth century (the Civil War and the abolition of slavery). No such vernacular evolved on the tobacco and cotton plantations, where the Africans were generally in the minority and were not segregated from the European majority until the late nineteenth century. Geographical and social isolation from mainland America until the mid-twentieth century helped preserve Gullah to this day as quite distinct from other American offspring of English.

It took coastal South Carolina colonists 30 years or so from the foundation of the colony in 1670 to shift from the homestead type of residence, characterized by no race segregation, to a partial plantation society, in which segregation was institutionalized as early as the mid-eighteenth century. (Virginia, which was founded in 1607, had not shifted to a largely agricultural economy, with tobacco plantations, until the end of the seventeenth century, either.) It also appears that, except for coastal Georgia (which patterned its economic and social development on South Carolina), in the rest of the southeastern English ex-colonies segregation was not institutionalized on a large scale until the late nineteenth century, after the passage of the Jim Crow laws.

This difference in the timing of segregation explains partly why Gullah has developed more non-standard characteristics than AAVE. It had a separate evolution, in which African languages exercised a greater influence on the integration of English materials into the emerging dialect. With a few exceptions, the vast majority of structural peculiarities associated with AAVE are shared by other American dialects, and the most significant differences between black and white forms of American English are statistical. In contrast, few of the features associated with Gullah are attested in other North American English vernaculars, although the features have definite English origins too. In areas inland from the coast, segregation was institutionalized after over two centuries of intimate coexistence of (descendants of) Africans and low-class Europeans, after many structural features had become entrenched in both their communities' dialects. Later changes associated today with differences between them remain minimal.

The development of vacation resorts in coastal Georgia and South Carolina and the urbanization of some of the Sea Islands has fostered residential patterns similar to those that have preserved AAVE in mainland cities. The affluent in-migrants, who are predominantly white, have settled in their own parts of the Islands. They have not interacted regularly with the Gullah populations who preceded them there by two and a half centuries.

There is arguably a geographical and structural continuum in African American speech from the South Carolina and Georgia coast through the mainland Southeast. Accordingly, Gullah represents the variety that is the most different from mainland white vernaculars. After all, although their sentence melodies are clearly distinct, Gullah and AAVE share quite a few features, such as constructions without the copula *be* (e.g. *she Ø pretty* 'she [is] pretty'); use of the negative *ain(t)* not only to mean 'be not' (e.g. *he ain crazy*) but also 'have not' or 'did not' (e.g. *he ain go/gone there* 'he has not gone there', 'he did not go there'); habitual *be* (e.g. *they be havin company* 'they are/were USU-ALLY IN THE PROCESS of having company'); and emphatic PERFECT with *done* (e.g. *she done gone* 'she has/had ALREADY gone'). (This is not to say that these peculiarities definitely did not grow out of English usages; but they have clustered in a way that makes African American speech distinct from other American nonstandard vernaculars.)

The geographical continuum hypothesis does not apply to Gullah alone, but also to White Coastal Plantation English, which is distinct from American White Southern English. One could thus also argue that Southeastern American English vernaculars are basically plantation varieties. Differences among them

have resulted from the kind of plantations on which they developed and the kinds of social contacts that were obtained across race boundaries on these plantations. The coastal vernaculars are rice-field phenomena, distinct from those that developed on the tobacco and cotton plantations.

From one evolutionary point of view, Gullah is considered the most extensively restructured variety of North American English (i.e. it shows the greatest influence of its creators' ancestral languages). However, this judgment reflects a traditional bias against all colonial vernaculars that took their vocabulary from European languages and were appropriated by descendants of Africans. It also reveals how little we yet know about vernaculars such as Old Amish English and Cajun Vernacular English, which developed from appropriations of English by isolated groups of proletarian continental Europeans. The opinion is also undermined by the absence of a reliable yardstick for assessing the extent of restructuring. Every North American English variety has developed from the contact of English dialects with each other (hence they too are restructured varieties) and with other languages, especially languages from Europe.

This restructuring-through-contact puts other American English vernaculars in the same category as Gullah and AAVE. It makes the often-invoked argument of mutual intelligibility (according to which Gullah is not intelligible to most European Americans) only one-sided and largely a reflection of the fact that Gullah developed from contact with languages that happened not to be European. There is little wondering about why the varieties that developed from the contact of English with other Western European languages (before the massive nineteenth-century immigrations from Central and Eastern Europe) are to some extent mutually intelligible, even though familiarity with a particular variety is an important factor in all such claims. Contacts of typologically similar varieties in similar ethnographic settings produced similar dialects, although listeners' attitudes toward particular speakers affect their ability to understand them. Future research may very well reveal that Gullah holds the most retentions from colonial English, although not necessarily in their pristine forms. An important factor in comparing vernaculars is that their features often originated in different English varieties. Neither they nor Gullah, as systems, could be claimed to be comprehensive retentions of any particular British English dialect.

There are no records to show how distinct Gullah was from AAVE before the nineteenth century. However, there is literary documentation of undifferentiated, stereotypical black speech, with pidgin-like features, since the early eighteenth century. This sug-

gests that a distinct variety of English was spoken by some African slaves in English colonies, but it does not confirm that Gullah as a distinct ethnic variety had already developed. Advertisements concerning runaway slaves in the mid-eighteenth century reveal that English among the African slaves varied from varieties close to those spoken by white colonists to others that were quite different, which may be assumed to be the beginnings of today's Gullah. Reports of 'fluent' or 'acceptable' command of colonial English generally correlated with whether the slave was American-born, was imported young to the colony, or, if imported as an adult, had lived here long enough. The absence of literary or other records of a divergent speech of Africans in the seventeenth century can be interpreted to suggest that no singularly African American vernacular was evident then. The generally nonsegregated living conditions of the time, with the colonial populations growing more by birth than by importation, lead to this conclusion.

The lack of reports of divergent speech suggests in fact that Gullah must have followed an evolutionary path similar to those of other colonial English vernaculars, differing from them essentially in the details of the restructuring process. The differences can be correlated to the following factors: The homestead phase during which the nonnative, non-European speakers were in the minority did not last long, only about 30 years after the founding of the South Carolina colony in 1670. This is a period during which the Africans were integrated into small family farm communities and interacted regularly with Europeans. Both European American and African American children spoke the same colonial koiné as vernaculars. By 1720, the population of African descent was double the population of European descent, and on the emergent plantations on the coast, the Africans became the overwhelming majority, often at a ratio of 9 to 1 by the late eighteenth century. The early institutionalization of race segregation forced the Africans to socialize primarily among themselves. The shift to rice cultivation as the primary colonial industry brought about some demographic changes, chiefly rapid population growth by importation of new labor while the mortality rate was high. Increasingly, language was transmitted to newcomers by nonnative 'seasoned' slaves, a factor that enabled more restructuring under dominant influence from African languages. This evolution may account for why Gullah has certain grammatical characteristics not found in other American dialects, including AAVE. As noted above, the development of Gullah must have started in the eighteenth century. However, it is not clear what its actual grammatical system was like then, especially how much it diverged from other colonial English vernaculars.

The features that distinguish Gullah from AAVE are not necessarily of non-English origin. For instance, the HABITUAL marker [dəz] (spelled *duhz*) is well attested in some British nonstandard vernaculars and in Newfoundland Vernacular English. Even the fact that Gullah has the option of expressing the PROGRESSIVE with [də] before the verb, and with *a* when the verb also has the ANTERIOR marker *bin* (as in *he bina talk* 'he was talking'), does not necessarily suggest that these characteristics stem from African languages. While African languages must have exerted some influence in the selection of specific particles that occur before the verb to mark tense and aspect, the forms themselves have typically been inherited from English.

It seems very helpful indeed to make a distinction in scholarly debate between the origins of features and the influence that determined their integration into the developing dialect. This is more obvious in Gullah's sound system, where one comes across sporadic uses of [ß] as a variant of [v] and [w], as in [ßɛri ßɛl] 'very well', which also alternates with [vɛri wɛl]. Typologically, the bilabial fricative is highly unusual in the sound systems of sub-Saharan African languages. [ß] is not attested in Caribbean English creoles, where African influence would be more extensive, nor in Caribbean French creoles, which had similar influences. Its counterpart in Caribbean English creoles is [b], as in [bɛks] 'angry' < English [vɛkst]. This suggests not necessarily the absence of [v] in some African languages, but primarily the inability of those who made the creoles to reproduce [ß], which is attested in Irish English and must have occurred in colonial English. The fact that Gullah has [ß] in alternation with the English variants [v] and [w], just like the fact that it has the schwa (unlike Caribbean English creoles), reflects peculiarities of the sociohistorical conditions under which it evolved.

African influence can, on the other hand, be identified in serial verb constructions, in the associative noun plural *an dem*, as in *Easter (an) dem* 'Easter and her family/friends/associates', and in some other constructions that it shares with AAVE and other creoles. Although serial verb constructions are more diversified in Gullah, these features are also attested in other American and British nonstandard vernaculars. However, their statistical predominance and some of their usage patterns reflect the influence that must have been exerted by some African languages during the development of Gullah.

References

Cunningham, Irma Aloyce Ewing. 1992. A syntactic analysis of Sea Island Creole. Tuscaloosa: University of Alabama Press.

Goodwine, Marquetta, and the Clarity Press Gullah Project. 1998. The legacy of Ibo landing: Gullah roots of African-American culture. Atlanta: Clarity Press, Inc.

Jones-Jackson, Patricia. When roots die: endangered traditions on the Sea Islands. Athens: University of Georgia Press.

Montgomery, Michael (ed.) 1994. The crucible of Carolina: essays in the development of Gullah language and cutlure. Athens: University of Georgia Press.

Pollitzer, William S. 1999. The Gullah people and their African heritage. Athens: University of Georgia Press.

Turner, Lorenzo. 1949. Africanisms in the Gullah dialect. Chicago: University of Chicago Press. Reprinted in 2002 with a new introduction by Katherine Wily Mille and Michael B. Montgomery. Columbia: University of South Carolina Press.

SALIKOKO MUFWENE

See also **African American Vernacular English; Pidgins and Creoles**

Gumperz, John Joseph

A blend of scientist, anthropologist, and linguist, John Joseph Gumperz's background is manifested both in his theoretical and applied research.

His research focused mainly on the complex nature of intercultural and interethnic communication or what came to be known as *interactional sociolinguistics*. Contrary to structuralists, who presuppose that human communication takes place in an ideal and nonproblematic context where there is almost no risk of misunderstanding, Gumperz gives substantial importance to the five elements involved in any verbal communicative event (the speaker, the listener, the context, the message, and the channel) using the macrosociological factors to interpret the microconversational effects. After his fieldwork in India, he shifted from purely dialectologist research to conversational analysis, yet without overlooking his early research interests. One of his most influential books, *Discourse strategies*, which appeared in 1982, is a product of both theoretical research in conversational analysis and his long years of fieldwork in India, Europe, and the United States, where he scrutinized both interracial and interethnic conversations. His main concern was with how the participants in a certain conversation behave according to their interpretation of the communicative intent. To answer this concern, he developed the theory of '*contextualization cues*', which he defines in *Discourse strategies* (p. 131) as 'any features of linguistic form that contributes to the signaling of contextual presuppositions'. These features may be prosodic, morphological, syntactic, lexical, or phonological. They may include other conversation behaviors such as turn-taking or code-switching. In the majority of the cases, these cues indicate how the speakers understand every part of the talk as being sequentially connected to both what has been said before and what is going to be said after. When the speakers fail to grasp and use the contextualization cues appropriately, misunderstanding could result and in some cases the communication line may break down even if the speakers shared other linguistic features. Learning another language does not guarantee effective communication unless the speakers are able to interpret the contextualization cues appropriately to keep redefining the context. Contextualization cues, then, are interpretive cues that guide the speakers in their inferential effort in any communicative event similar, in this sense, to the Gricean Maxims of conversation (Grice 1975). While the Gricean Maxims are theoretically applicable to universal communication, the contextualization cues may exist across cultures but they are still partly culture-bound and even conversation-bound.

Gumperz's interpretation of *code-switching* as a contextualization cue has been very significant in shaping later research on this complex sociolinguistic phenomenon. He claims the existence of two types of code-switching: *situational code-switching* and *metaphorical code-switching*. The first is produced when the codes used by the speakers change according to the redefinition of the situation in which they are involved. A common example is when a monolingual speaker joins a conversation already started by bilingual speakers; they would switch to his or her language so as not to exclude him or her. Metaphorical code-switching, on the other hand, is caused by the change in the topic of the conversation without changing the situation. In the following example (Gumperz 1982a:77), the switch to Hindi by speaker **B** shows the rupture with the formal conversation that he was carrying out with speaker **A**. The change in the topic of the conversation stands behind the switch and not the change in the speakers since speaker **C** was present from the beginning.

'A group of Hindi speaking graduate students are discussing the subject of Hindi–English code-switching:

A: Sometimes you get excited and then you speak in Hindi, then again you go on to English.

B: No nonsense, it depends on your control of English.

B: [shortly thereafter turning to a third participant, who has just returned from answering the doorbell] **Kɔn hai bai** (who is it)?'

One of the main explanations for the occurrence of code-switching is the *in-group* vs. *out-group* theory as elaborated by Gumperz (1972, 1982a). He refers to code-switchers as the individual speakers who, for different reasons, find themselves living in an ethnically and culturally diverse setting. This includes the case of immigrants who usually manage to keep their first language (L1) to communicate among themselves in their restricted speech communities, but who find themselves using the majority's language, generally the official language of the host country, basically at work and education. Within the boundaries of the '*in-group*', the individual speaker tends to use the '*we code*', i.e. the L1 of his or her own speech community. An example is Spanish for the Puerto Rican community in New York extensively studied by Gumperz and others. During the interaction with members of the '*out-group*', i.e. members of the majority or any other speech community, the tendency is to use the '*they code*', which is the majority language.

In some cases, the continuous switch back and forth between two codes can be an indicating strategy of the speaker's neutrality with regard to contending parties or ideologies that are associated with different codes. On the other hand, religious or political issues, among others, are often discussed in the code that is most often associated with them.

In his later book edited with S.C. Levinson, *Rethinking linguistic relativity*, Gumperz revisits his theory of contextualization cues to comment on the relationship that language has with thought and culture and the way it shapes our conception of the world. The editors signal the changes in the understanding of the Sapir-Whorf hypothesis and its implication for the theories of linguistic relativity. Here again, Gumperz manages to give equal importance both to the macrosociological and cultural factors and the very specific microconversational features in each communicative event.

Biography

Born in Germany in 1922, Gumperz moved to the United States in 1939 where he received a Bachelor of Science from the University of Cincinnati. He later earned a Ph.D. in German linguistics from the University of Michigan. In 1955, he conducted fieldwork in India. And in 1965, he was appointed as an anthropology professor at Berkeley where he is still an emeritus Professor today.

Since the appearance of his first article early in 1955 ('The phonology of North Indian Village Dialect'), Gumperz has contributed consistently to different publications (Berkeley Cognitive Science Report, The American Journal of Sociology, American Speech, etc.). In addition he has written and co-edited several other influential books including The Ethnography of Communication of (1964), Directions in Sociolinguistics (1972), and Language and Social Identity (1982).

References

Grice, H.P. 1975. Logic and conversation. Syntax and semantics: Speech acts, ed. by In P. Cole and J.L. Morgan (eds). Volume 3, 41–55. New York: Academic Press.

Gumperz, John Joseph. 1955. The phonology of a North Indian village dialect. Indian Linguistics 16. 283–95.

_____. 1964. The ethnography of communication (co-edited with Dell Hymes). American Anthropologist 66(6, part 2, Special Issue).

Gumperz, John Joseph, and Dell Hymes (eds.) 1972. Directions in sociolinguistics. New York: Holt, Rinehart, and Winston.

_____. 1972. Social meaning in linguistic structures: code switching in Northern Norway(cowritten with Jan-Petter Blom). Directions in sociolinguistics, ed. by J.J. Gumperz and D. Hymes. New York: Holt, Rinehart, and Winston.

_____. 1982a. Discourse strategies. Cambridge: Cambridge University Press.

_____. 1982b. Language and social identity. Cambridge: Cambridge University Press.

_____. 1984. Communication and social identity. Identity: personal and socio-cultural, ed. by A. Jacobson-Widding. Uppsala Studies in Cultural Anthropology, Vol. 5.

_____. 1990. Contextualization and understanding. Rethinking context, ed. by A. Duranti and C. Goodwin. New York: Cambridge University Press.

_____. 1996. Rethinking linguistic relativity (co-edited with Stephen C. Levinson). Cambridge: Cambridge University Press.

LOTFI SAYAHI

H

Haas, Mary Rosamond

The life and growth of language: *an outline of linguistic science* (1875) by William Dwight Whitney introduced Mary Haas to linguistics, and in 1930, following her undergraduate studies at Earlham College, she set off for the University of Chicago intending to study comparative philology. A class in historical Germanic linguistics taught in German by Leonard Bloomfield did not appeal to her, but when she ventured into the anthropology department where Edward Sapir was elaborating upon the topics he had covered in his book *Language* (1921), she encountered the kind of linguistics to which she would devote much of her career.

She trained as an Americanist, a term used by anthropologists and anthropological linguists for the study of the native languages and cultures of North America. Her first fieldwork was undertaken with her then-husband Morris Swadesh. With support and direction from Sapir, the couple traveled to Vancouver Island in Canada in the summer of 1931 to collect texts and data on Nitinat. Their joint report was Mary Haas's first publication.

In Fall 1931, Sapir joined the faculty at Yale University. Haas, Swadesh, and several other graduate students followed him, becoming the nucleus of the 'Sapir School' of linguistics, focused on the collection of texts and the preparation of descriptive grammars of native American languages, with a strong concurrent interest in the extension of concepts and techniques from comparative Indo-European linguistics to the historical and comparative study of American languages.

As a graduate student, Haas worked on Tunica, a language spoken in Louisiana. 'A Grammar of the Tunica Language' became her 1935 Yale doctoral dissertation; it was later published with revisions in Franz Boas's *Handbook of American Indian languages* (1941) and remains an outstanding exemplar of Sapirian anthropological linguistics. Meanwhile, Haas spent her postdoctoral years as a fieldworker, gathering data on three Muskogean languages, Creek, Koasati, and Hichiti, as well as on Natchez. Her classic study of gender and language, 'Men's and women's speech in Koasati', appeared in 1944 in *Language*, journal of the Linguistic Society of America. Other studies drawn from the data she had collected were published later in her career; much remained unpublished.

World War II brought a new opportunity for paid work, and Haas became the first linguist supported by the Intensive Language Program of the American Council of Learned Societies (ACLS). The goal here was to develop descriptions and teaching materials for lesser known languages of Europe and Asia. She was assigned to work on Thai and when she located native speakers of the language studying at the University of Michigan, she moved there to work with them. Her article 'The linguist as a teacher of languages' (*Language* 1943) discussed how she applied her descriptive linguistic techniques to the development of materials and classroom experiences in foreign language learning. Haas went on to publish more than 20 articles and books on the Thai language, including a textbook and a reader prepared with her second husband, Heng R. Subhanka. Her analyses of various aspects of the Thai language, including its phonology, morphology, and writing system, are notable for their clarity and insight and remain valuable descriptive resources even today.

Following a brief ACLS-sponsored pilot class in Thai at the University of Michigan, Haas moved to the University of California, Berkeley as a lecturer in what was there termed Siamese. She also taught linguistics courses and was instrumental in the development of the Berkeley graduate department of linguistics. In 1953, her colleague Murray B. Emeneau obtained funding for the establishment of the Survey of California Indian Languages, and Mary Haas was appointed as director. This program provided the research center of the department, with graduate students expected to focus their long-term work on a single language, writing a descriptive grammar for a dissertation and preparing a dictionary and set of texts for publication, often in the University of California Publications in Linguistics series.

With the first half of her career devoted to the documentation of languages, Haas returned in the second half to the historical comparison of languages that had initially drawn her to the field. A lengthy 1966 article on 'Historical linguistics and the genetic relationship of languages' became *The prehistory of languages* (1969), her only book on a general linguistic topic. Here, she used her earlier fieldwork on the Muskogean language family, along with data from other native American languages, to illustrate types of language change, problems in genetic classification, and the reconstruction of protolanguages.

Throughout her career, Haas remained steadfast in her commitment to the model of Americanist linguistics developed originally by Franz Boas and then by Edward Sapir. She rejected what she felt was the rigidity of post-Bloomfieldian structuralism and was openly dismissive of some of the linguistic work of her east coast contemporaries. In the later 1950s and thereafter, Haas virtually ignored the developments in generative grammar that dominated American linguistics. She had little interest in studies drawing on English or the languages of Europe, and she believed that linguistic fieldwork was threatened by the use of their own language by generative linguists who relied on their native knowledge for linguistic insights. Haas continued on her own path, documenting the languages of native America, comparing them, and proposing genetic classifications. Toward the end of her career, she began to look back at the history of her specialization. A set of articles on the history of American Indian linguistics is reprinted in a collection of her essays (*Language culture, and history* 1978) and remains a valuable account, sometimes firsthand, of how that field developed in the United States.

Mary Haas was not interested in linguistic theory, but she was fascinated by the data of languages and she conveyed this to her students. Her most lasting legacy is the cadre of some 40 graduate students who, under her direction or with her assistance in the 1950s, 1960s, and 1970s, earned doctorates with work on American Indian languages. Not particularly effective as a classroom teacher, Haas was an inspiring mentor. Her students and their students can still be found in the linguistics departments of American and Canadian colleges and universities, many of them carrying on her Americanist tradition.

Biography

Mary Haas was born in Richmond, Indiana, on January 23, 1910. She earned the AB at Earlham College in 1930 and began her graduate study at University of Chicago in 1930–31. She carried out fieldwork on Nitinat in the summer of 1931 with Morris Swadesh. She went to Yale University in the fall of 1931, and completed her Ph.D. in 1935 with a dissertation on Tunica. She pursued postdoctoral independent research on Muskogean languages and Natchez. In 1941–1943, she pursued research, materials preparation, and pilot teaching of Thai at the University of Michigan under the Intensive Language Program of the American Council of Learned Societies. In 1943–1944 and 1946–1947, she was a lecturer in Siamese (Thai) at the University of California, Berkeley, and in 1947–1952, she was assistant professor of Siamese and linguistics. In 1952–1957, she was associate professor; in 1957–1977, professor of linguistics; department chair in 1958–1964; and director Survey of California Indian Languages (later called Survey of California and Other Indian Languages) in 1953–1977. She retired in 1977. She was visiting professor or fellow at University of Washington, Ohio State University, Barnard College and Columbia University, Australian National University, Northwestern University, University of California, San Diego, University of Georgia, University of Kansas and University of Calgary. She was also Linguistic Society of America vice president in 1956 and president in 1963. She was Guggenheim Fellow in 1964–1965 and concurrently fellow of the National Endowment for the Humanities and the Center for Advanced Studies in the Behavioral Sciences in 1967–1968; she was also elected to National Academy of Sciences in 1978. She received honorary degrees from Northwestern University in 1975, University of Chicago in 1976, and Earlham College in 1980. Haas died in Berkeley, California on May 17, 1996.

References

Campbell, Lyle. 1997. American Indian languages: the historical linguistics of native America. New York and Oxford: Oxford University Press.

Goddard, Ives. (ed.) 1996. Languages, Handbook of North American Indians, Vol. 17, series editor William C. Sturtevant. Washington, DC: Smithsonian Institution.

Golla, Victor, James A. Matisoff, and Pamela Munro. 1997. Mary R. Haas. Language 73. 826–37.

Haas, Mary Rosamond. 1932. A visit to the other world, a Nitinat text (with translation and grammatical analysis), as Mary Haas Swadesh with Morris Swadesh. International Journal of American Linguistics 7.

———. Tunica. 1941. Handbook of American Indian languages, Vol. 4, ed. by Franz Boas. New York: J. J. Augustin.

Haas, Mary Rosamond, and Heng R. Subhanka. 1945, 1948. Spoken Thai, 2 vols., 1945 (Book I), 1948 (Book II). New York: Henry Holt.

———. 1956. The Thai system of writing, Washington, DC: American Council of Learned Societies.

Haas, Mary Rosamond, with George Grekoff, Ruchira C. Mendiones, Waiwit Buddhari, Joseph R. Cooke, and Søren C. Egerod. 1964. Thai–English student's dictionary, Stanford, CA: Stanford University Press.

Haas, Mary Rosamond. 1969. The prehistory of languages. The Hague and Paris: Mouton.

The following articles, as well as many others, are reprinted in Language, culture, and history: essays by Mary R. Haas, selected and introduced by Anwar S. Dil. Stanford, CA: Stanford University Press (1978):

The linguist as a teacher of languages. Language 19 (1943)

Men's and women's speech in Koasati. Language 20 (1944)

Historical linguistics and the genetic relationship of languages. Current trends in linguistics, Vol. 3, ed. by Thomas A. Sebeok. The Hague and Paris: Mouton (1966).

Hinton, Leanne. 1993. Women in southwestern linguistic studies. Hidden scholars: women anthropologists and the native American Southwest, ed. by Nancy J. Parezo. Albuquerque: University of New Mexico Press.

Hinton, Leanne.(ed.) 1997. The Hokan, Penutian, and J. P. Harrington Conferences and the Mary R. Haas Memorial, 28–29 June 1996, University of California, Berkeley (Survey of California and Other Indian Languages Report 10), Berkeley: Department of Linguistics, University of California.

Parks, Douglas R. (ed.) 1997. Mary R. Haas: a memorial issue: a collection of essays. Anthropological Linguistics 39.

Shipley, William. (ed.) 1988. In honor of Mary Haas: from the Haas Festival Conference on Native American Linguistics. Berlin, New York, and Amsterdam: Mouton de Gruyter.

JULIA S. FALK

See also **Sapir, Edward; Swadesh, Morris; Thai and Tai Languages**

Haitian Creole

Haitian Creole is one of the many French-based Creoles that were established throughout the world during France's colonizing period. Creole languages are commonly accepted as being a language that was developed from an earlier, simpler version of that language, called a Pidgin.

Pidgins arose in situations of linguistic need (such as for trade or other economic purposes) between two or more groups of people who spoke mutually unintelligible languages. After a generation or two, the offspring of the original Pidgin speakers tended to develop a more complex language, with a wider range of expressive and social functions. These languages are referred to as Creoles, and their vocabulary is generally based on the lexicon of the colonizing group.

Although Haitian Creole is French-based, it also contains elements of African influences from Wolof, Fon, and Ewe. The basic word order is Subject, Verb, Object (SVO), but typical Creole markers appear in the use of postpositions, articles, and possessives coming after the noun (e.g. *nom-la*, 'man' 'the').

Two theories have been proposed to explain the ultimate origin of Creole languages. The claim for the monogenetic theory is that Creoles all over the world have a common origin; a Portuguese-based pidgin was created in Africa in the fifteenth century. During the establishment of the Creole, the Portuguese and native words were relexified or changed into the lexical items of the colonizing or *superstrate* language. The polygenesis theory suggests that whenever and wherever a contact situation arose between two or more languages, a pidgin was created, which subsequently led to the establishment of a Creole. This theory, however, fails to account for the amazing amount of similarity between Creole languages in widely disparate geographical areas of the world. In addition, Bickerton (1981) promotes the 'bioprogram' as the basis for the genesis of Creoles. Building on the idea of language universals and language simplification processes common to first-language acquisition, Bickerton suggested that humans are genetically programmed to acquire a specific set of linguistic structures, and that these structures are modified later in life as children are exposed to their native language. The commonality among these theories is the acceptance that the superstrate language (French, English, Spanish, etc.) provided the lexical base, while the syntactic elements were organized according to the *substrate* or original language used by Creole speakers.

For the origin of Haitian Creole French, Valdman (1997) claims that the earliest form of a French-based creole appeared first on the island of St. Kitts; this language was subsequently exported to Guadeloupe and Martinique, and then to Haiti. He extrapolates from this diffusion theory to claim that the newest, most

innovative form of a Creole language will be on those islands from which it first came—Martinique, Guadeloupe, and Haiti—and that areas that were later recipients of the language (such as Louisiana) will exhibit more conservative forms of the language.

One reason for language change (from conservation to innovation, for example) is due to the process of *decreolization*. This occurs when the use of the standard form of the superstrate is viewed as prestigious, and the use of the Creole is given low status. Because of the history of colonization and slavery often associated with Creole languages and the people who speak them, this low prestige is exacerbated, and causes a shift toward the use of the standard form of the language. This results in a continuum of varieties: the *basilect*, which is farthest from the superstrate language, the *mesolect*, an intermediate variety between the two, and the *acrolect*, the code closest to the substrate language. This *postcreole continuum* exists in Haiti, and, just as in many areas of the world where Creoles are spoken, speakers often have control over several of the varieties. When and how they use them is often dependent upon social factors such as age, gender, ethnicity, and geographical origins of the speaker and hearer. In addition, identification as Haitian often drives the use of the Creole rather than a more Standard French code. Haitians take much pride in the fact that their island was the site of the first slave revolt in modern history (1791–1804), which abolished slavery and established Haiti as the second independent nation in the Americas (Hall 1992). As a result, they have clung tenaciously to their (Creole) language, their history, and their culture.

This linguistic strength was shown in 1961 when Haitian French was granted legal and educational status in Haiti. It was accepted as the official language of the island in 1987 (along with Standard French), and has since developed a standardized written code and a growing literature. There are several dictionaries of Haitian Creole, and newspapers, articles, fiction, and poetry are being written in the language. Radio and television broadcasts are in Haitian Creole, and there was a bible printed in the language in 1985. Although economic conditions are abysmal, and political upheavals are the norm, the language appears to have been cemented in the local conscience, and therefore plays an important role in identification as Haitian.

Haitian Creole is spoken on the western portion of the island of Hispaniola; the other part of the island is the Spanish-owned Dominican Republic. There are estimated to be over five million Haitian Creole speakers at present, who represent about 90% of the island population, and constitute the largest number of Creole speakers anywhere. In addition, there are many substantial Haitian enclaves in the Bahamas, Canada, and the United States. Ninety-five percent of Haitians are of African origin, but the extremely small mulatto (mixed-race) and French population constitutes the upper class. This is also the group that has continued to use and speak standard French. Recent language attitude studies and the political climate of the island suggest that linguistic, social, and cultural mores are moving in a direction of viewing Creole as a positive force, while simultaneously rejecting or lowering the status of the standard French language and culture, which was imposed on the island so many centuries ago.

References

Bickerton, Derek. 1981. Roots of language. Ann Arbor: Karoma.

Ferguson, Charles. 1957. Diglossia. Word 15. 325–40.

Holm, John. 1988. Pidgins and Creoles, vols. 1 and 2. Cambridge: Cambridge University Press.

Hymes, Dell. (ed.) 1971. Pidginization and creolization of languages. Cambridge: Cambridge University Press.

Lefebvre, Claire, Hélène Magloire-Holly, and Nanie Piou. (eds.) 1982. Syntaxe de l'haïtien. Ann Arbor: Karoma.

Le Page, R.B. (ed.) 1961. Creole language studies II: Proceedings of the Conference on Creole Language Studies, University of the West Indies, Mona, 1959. London: Macmillan.

Mühlhäusler, Peter. 1986. Pidgin and Creole linguistics. Oxford: Basil Blackwell.

Romaine, Suzanne. 1988. Pidgin and Creole languages. London: Longman.

Valdman, Albert. 1977. Pidgin and Creole linguistics. Bloomington: Indiana University Press.

Valdman, Albert. 1978. Le créole: structure, statut et origine. Paris: Klincksieck

<div align="right">MEGAN E. MELANÇON</div>

See also **Pidgins and Creoles**

Hale, Kenneth

Like his teacher Carl Voegelin, Kenneth Hale was a prolific documenter and avid student of languages. He learnt Spanish in his youth in southern Arizona, and Navajo and O'odham (Papago). His training combined linguistics and anthropology in the Americanist tradition following Edward Sapir.

His perennial interests covered many areas of linguistics, including language typology, lexicography, syntactic theory, historical linguistics and language reconstruction, and applied linguistics in school education especially in support of minority languages.

Hale's early major work comprised several careful comparative studies of language groups based on data he collected himself, first Uto-Aztecan, then subgroups of Australian languages (Arandic, Northern Pamic). From his 1959–1960 fieldwork, Hale perceived that the phonologically aberrant languages of northern Cape York Peninsula were related by regular sound changes to other Australian languages; he further showed the implications of this for prehistory. Hale's 1961 coinage 'Pama-Nyungan' has since been generally applied to name the largest classificatory grouping of Australian languages, the outline of which appeared in 1966 (with Geoff O'Grady and the Voegelins). Hale's extensive notes, analyses, transcriptions, and sound recordings of about 70 Australian languages, archived at the Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) in Canberra, are a great accomplishment, and have provided a basis for much subsequent work by others.

Hale was inspired by Noam Chomsky's *Aspects of the theory of syntax* and arranged to return to Australia and reexamine some Australian languages (Warlpiri and Lardil) with this new framework. While remaining sensitive to intricate details of complex phenomena (such as pronominal marking in clauses) at the same time Hale believed that linguistic analysis needed to 'strip away complexity' and reveal simpler principles and modules at work in combination, relating the universal and the particular. Thus, Hale was able to bridge from his Sapirian training to Chomskyan universal grammar. One of Hale's influential proposals was a later reaction (1983) to the *Aspects* model, in which he argued, on the basis of his studies of languages such as Warlpiri and O'odham (Papago), that in some languages, the mapping from (logical) argument structure to syntactic configuration works quite differently than it does in English. Such 'nonconfigurational' languages do not reserve a special syntactic position for clause subject (or object), and are expected to exhibit a contrasting cluster of syntactic properties, notably clausal word order independent of syntactic relations. This subsumed another of Hale's related influential proposals, that there are languages (such as Warlpiri) where relative clauses are not embedded (as they are in languages like English) but rather simply added (adjoined) to their host clause. For the rest of his life, he continued to explore how conceptual argument structure relates to grammar, word order, and word structure, looking at languages

as diverse as Chukchi, Dagur, English, Hopi, Miskitu, Igbo, Irish, Japanese, Navajo, O'odham, Warlpiri, and West Greenlandic, often in collaboration with native speakers of the languages.

Hale's reputation as a polyglot was well-deserved, as he commanded dozens of languages. He had an extraordinary linguistic memory: he was able to absorb intensely not only the grammar and vocabulary of a language but also the precision of pronunciation. He downplayed this ability, since he believed that full knowledge of a language also included knowing its 'phrase book' with thousands of common expressions and their appropriate use. His annual class at MIT explored the grammar of one language each year, usually with a native-speaker linguist, and included sessions of monolingual investigation. Over 30 years, the class ranged from several North American languages, to the likes of Maori, Warlpiri, Finnish, Amoy, Malayalam, Basque, Tamazight Berber, Abkhaz, Zarma, Igbo, Yoruba, Korean, and American Sign Language (ASL).

In language investigation and description, Hale drew on his anthropology training in sensitivity to various linguistic registers (appropriate to gender, age, kin relations, ceremonial situations), and to the language of songs. He did not just draw on his own abilities, but showed how the native speaker could be a collaborator, and, in particular, he demonstrated the value and multiple uses of good 'folk definitions' or 'ethnosemantic oral essays'. Many other linguists enjoyed collaborating with him; he coauthored several works with each of Geoffrey O'Grady, LaVerne Masayesva Jeanne, Paul Platero, Samuel Jay Keyser, and Maria Bittner.

Hale campaigned for action to stem the endangerment of languages. Conscious that knowledge was empowering, he sought to put the results of his linguistic and anthropological research at the disposal of the speakers of those languages. He encouraged bilingual education, coauthoring a 1974 recommendation with Geoff O'Grady on which bilingual education programs in the Northern Territory of Australia were founded. He also helped speakers of local languages obtain training in linguistics both at workshops in their communities, and at MIT, so that they could study the language themselves and get some control of its destiny. From 1986, he made several visits to work with local people to study Misumalpan languages (Miskitu, Sumu) of the Atlantic Coast of Nicaragua and Honduras. He was instrumental in the development of dictionaries, including for Warlpiri, Lardil, and Ulwa and Miskitu (two Misumalpan languages native to Nicaragua), and during his last years, he mentored the Wampanoag revival project in Massachusetts.

Under Hale's direction in the 1980s, the Lexicon Project at the Center for Cognitive Science, MIT carried out comparative lexicology (research on the nature of the lexicon) with a view to discovering general lexico-semantic themes, or principles that recur systematically in the definitions of lexical items. The languages Hale himself focused on in this project were the Central Australian language Warlpiri, Miskitu and Southern Sumu (Ulwa), and English. Later, Hale's ideas on argument structure and the nature of thematic roles were the focus of his major book (with S.J. Keyser) *Prolegomenon to a theory of argument structure*, which was brought to completion as he succumbed to cancer.

Biography

Kenneth Hale was born in Evanston, Illinois on August 15, 1934; his family moved to Canelo, Arizona by 1941. In 1952–1955, he received his B.A. (1955) in Anthropology at University of Arizona, M.A. in 1956, with a thesis on *Class II Prefixes in Navajo*, and Ph.D. in l959, with a dissertation on *A Papago Grammar*, both at Indiana University, Bloomington, Indiana under Carl Voegelin. He was part of the linguistic survey of Australian languages conducted in 1959–1960, affiliated with University of Sydney, part of CF and FM Voegelin's survey of Languages of the World. He was Assistant Professor in Anthropology and Linguistics, University of Illinois, Urbana, Illinois in 1961–1964 and University of Alberta, Edmonton, in the summer of 1964, and was also Associate Professor in Anthropology, University of Arizona, Tucson, in 1964–1966. His second stay in Australia was in 1966–1967, again supported by National Science Foundation, focusing on Warlpiri and Lardil languages. He was recruited by Morris Halle in 1966 to Massachusetts Institute of Technology, where he was Associate Professor from 1967, and was promoted to Professor in 1972, and finally became Ferrari P. Ward Professor of Modern Languages and Linguistics in 1981–1999. His visiting appointments included University of Arizona 1976–1977; Katholieke Hogeschool Tilburg, Netherlands, 1983–1984; Australian Linguistic Institute, La Trobe University, Melbourne, July 1994; Linguistic Society of America Linguistic Institutes, University of Arizona, Tucson, in the summer of 1989; and University of New Mexico, Albuquerque, in the summer of 1995.

He was active in training native speakers of indigenous and minority languages: Warlpiri (Yuendumu School, Central Australia, mid-1974); Navajo (Kin Li Chee, Arizona, summer 1975, Navajo Community College, Tsaile, Arizona, summer 1982); in Miskitu and Sumu (Programa Educativo Bilingüe-Bicultural, Puerto Cabezas, Zelaya Norte, Nicaragua, January l986 and 1994, and Bluefields, Region Autónoma Atlántico Sur, Nicaragua, January, 1988 and 1990); and Mayan languages (Proyecto Lingüístico Francisco Marroquín, Antigua Guatemala, July 1988, and CIRMA, Antigua Guatemala, August 1994.

He was a member of the Linguistic Society of America (LSA), and its President in 1994–1995. He was also a member of the American Academy of Arts and Sciences from 1989, and of the National Academy of Sciences from 1990. Hale died in Lexington, Massachusetts on October 8, 2001.

References

Carnie, Andrew, Eloise Jelinek, and Mary Ann Willie. (eds.) 2000. Papers in honor of Ken Hale. Working papers in endangered and less familiar languages, Vol. 1. Cambridge, MA: MIT working papers in linguistics.

Hale, Keneth and Carl F. Voegelin, and Florence M. Voegelin. 1962. Typological and comparative grammar of Uto-Aztecan: I (Phonology). Indiana University Publications in Anthropology and Linguistics, Memoir 17 of the International Journal of American Linguistics.

Hale, Kenneth. 1964. Classification of Northern Paman languages, Cape York Peninsula, Australia: Research Report. Oceanic Linguistics 3(2).

Hale, Kenneth, and Joseph B. Casagrande. 1967. Semantic relationships in Papago folk-definitions, Studies in Southwestern ethnolinguistics: meaning and history in the languages of the American Southwest, ed. by D. Hymes and W. Bittle. The Hague: Mouton.

———. 1972. Some questions about anthropological linguistics: the role of native knowledge Reinventing anthropology, ed. by Dell Hymes, 382–97. New York: Pantheon.

———.1973. A note on subject–object inversion in Navajo. Issues in linguistics: papers in honor of Henry and Renée Kahane, ed. by B.B. Kachru et al., 300–9. Urbana: University of Illinois Press.

———. 1973. Deep surface canonical disparities in relation to analysis and change: an Australian example. Current trends in linguistics, ed. by T.A. Sebeok, 401–58. The Hague: Mouton.

———.1976. The adjoined relative clause in Australia. Grammatical categories in Australian languages, ed. by R.M.W. Dixon, 78–105. Canberra: Australian Institute of Aboriginal Studies, and New Jersey: Humanities Press.

———.1976. Linguistic Autonomy and the Linguistics of Carl Voegelin. Anthropological Linguistics l8.

———.1982. Some essential features of Warlpiri verbal clauses. Papers in Warlpiri grammar: in memory of Lothar Jagst, ed. by Stephen Swartz, 217–315. Work Papers of SIL-AAB, Series A, Vol. 6. Berrimah, NT: SIL-AAB.

Hale, Kenneth.1983. Warlpiri and the grammar of non-configurational languages. Natural Language and Linguistic Theory 1(1). 5–47.

———.1986. Notes on world view and semantic categories: some warlpiri examples. Features and projections, ed. by Muysken and H. van Riemsdijk, 233–54. Studies in Generative Grammar, Vol. 25. Dordrecht: Foris.

———.1992. Subject obviation, switch reference, and control. Control and grammar, ed. by R.K. Larson, S. Iatridou,

U. Lahiri and J. Higginbotham, 51–77, Studies in Linguistics and Philosophy, Vol. 48. Dordrecht: Kluwer.

Hale, Kenneth, and Albert Alvarez. 1970. Toward a manual of Papago grammar: some phonological terms. International Journal of American Linguistics (Memorial Issue for HansWolff, Fascicle II) 36(2).

Hale, Kenneth, Colette Craig, Nora England, Laverne Masayesva Jeanne, Michael Krauss, Lucille Watahomigie, and Akira Yamamoto. 1992. Endangered languages. Language 68(1). 1–42.

Hale, Kenneth, and Leanne Hinton (eds.) 2001. The Green Book of language revitalization in practice. San Diego: Academic Press.

Hale, Kenneth, and Samuel Jay Keyser. 2002. Prolegomenon to a theory of argument structure. Linguistic Inquiry Monograph 39. Cambridge, MA: MIT Press.

Kenstowicz, Michael (ed). 2001. Ken Hale: a life in language. Cambridge, MA: MIT Press.

Keyser, Samuel Jay. 2003. Obituary: Kenneth Locke Hale. Language 79. 411–22.

Laughren, Mary. 2001. Obituary Kenneth Locke Hale (1934–2001). Aboriginal History 25. 265–75.

Simpson, Jane, David Nash, Mary Laughren, Peter Austin, and Barry Alpher (eds.) 2001. Forty years on: Ken Hale and Australian languages. Canberra: Pacific Linguistics 512.

Yengoyan, Aram. 2003. Kenneth L. Hale (1934–2001). American Anthropologist 105(1). 222–5.

DAVID NASH

See also **Chomsky, Noam; Configurationality; Sapir, Edward; Thematic Structure**

Halle, Morris

Morris Halle can be called the father of the modern study of speech sounds (i.e. phonological and phonetic theory). He was the central figure behind the development of Generative Phonology, and has had a significant impact on the study of word structure (morphology). In addition to his intellectual stature, Halle's influence as an educator is profound. He has taught and advised many of the leading figures in phonology today, including those who have proposed competing theories.

Halle began his study of structural phonology under Roman Jakobson, first at Columbia University, and later at Harvard. In his early work with Jakobson and in collaboration with Gunnar Fant, Halle started his career by questioning the idea that phonemes, i.e. sound units roughly corresponding to alphabet letters, function as the basic cognitive units in the processing of speech sounds. Jakobson et al. (1951), in the spirit of work done in the Prague School of linguistics, claimed that phonemes are themselves divisible into smaller units, so-called 'distinctive features', which capture articulatory and acoustic characteristics of speech sounds. For example, the distinction between the word *pat*, pronounced [pæt], and *bat*, pronounced [bæt], does not lie in some categorial distinction between the cognitive units /p/ and /b/, but in the fact that only the latter is pronounced with vibrating vocal cords. In later work, Halle and his students developed hierarchically organized models of feature systems to capture the fact that some features seem to be more prominent than others in the sound systems of specific languages.

While working on his dissertation, Halle began working with researchers at the Research Laboratory of Electronics (RLE) at the Massachusetts Institute of Technology. He later began teaching in the Modern Languages department there. In the mid-1950s, Halle was a driving force behind the move to develop a linguistics program at MIT. Part of building up this program included hiring Halle's collaborator and colleague Noam Chomsky. Taking a page from Chomsky's syntactic work, Halle claimed that phonological information is not merely perceptual discrimination and articulatory action, but is a cognitive system with its own set of principles. These principles form the foundation of Generative Phonology, so-called because it was assumed that all speech sounds could be *generated* on the basis of a finite set of principles or rules. The classic book of Generative Phonology is Chomsky and Halle's (1968) *The sound pattern of English* (SPE). Classical Generative Phonology assumes that there are two levels of phonological information: the underlying or base form of a word and the surface or phonetic form. The underlying form is mapped to the phonetic form by a series of ordered generative rules.

Consider, for example, the pronunciation of the plural suffix –*s* in the words *dogs* [dagz], *cats* [kæts], and *glasses* [glæsəz]. The underlying form of this plural suffix is posited to be /z/. The phonetic form of the noun–suffix combination is generated via two rules.

The first rule states that a consonant is devoiced, i.e. pronounced without vibrating vocal cords, when it follows a voiceless consonant. The second rule, called

epenthesis, inserts a schwa /ə/ sound between adjacent coronal sibalents.

The effect of these rules is seen in the following derivations of the words mentioned above:

underlying representation	/dag/ + /z/	/kæt/ + /z/	/glæs/ + /z/
epenthesis	——	——	glæsəz
devoicing	——	kæts	——
phonetic representation	[dagz]	[kæts]	[glæsəz]

It is argued that epenthesis must apply before the devoicing rule, since applying them in the other order would result in an incorrect form *[glasəs], as the final /s/ of *glass* is a voiceless sound, and the /z/ would be devoiced when adjacent to it:

underlying representation	/glæs/ + /z/
devoicing	glæss
epenthesis	glæsəs
phonetic representation	*[glæsəs]

Besides applying the theoretical framework of generative grammar to sound systems, *The Sound Pattern of English* also provides one of the most complete discussions of English phonology so far.

Metrical theory (or the study of stress) forms another important strand in Halle's research program. SPE contained the first serious description of English word stress in terms of predictable rules. Building upon the success of autosegmental phonology, in more recent years, Halle applied a notion of autosegmental structure to the study of stress. Throughout this research, Halle has not only examined stress by examining individual words, but also words in phrasal context and in verse.

In generative phonology, the sound system closely interacts with the structure of the containing words. Thus, Halle is an influential figure in the subdiscipline of morphology as well. By applying the familiar generative model to words, he developed the first explicit theory of generative morphology. This theory is a lexicalist theory of word formation (i.e. all the operations of word formation occur in the speaker's lexicon or mental dictionary). Generative morphology consisted primarily of a list of morphemes and a set of word formation rules (WFRs) which built words out of the morphemes. Work in generative morphology eventually evolved into the theory of Lexical Phonology.

In the 1990s, Halle moved away from the lexicalist position his work had previously taken, and developed the theory known as Distributed Morphology (DM), in collaboration with Alec Marantz. Distributed Morphology holds that words are not constructed in the lexicon. Instead, syntactic operations which influence morphological form can occur before the words are inserted. The theory is "distributed" because what was previously thought to occur in the lexicon is distributed into three grammatical systems: (i) a set of syntactic primitives, (ii) a list of phonological forms and the syntactic environments in which they occur and (iii) a list of sound-meaning pairs. The first system comes into play as the syntactic rules construct the sentence, the second inserts the words into the sentence and the third is the tie between the surface phonological form and the semantic structure of the word or phrase.

The influence that Morris Halle has had on the field of linguistics is extensive. He has personally trained an enormous number of today's leading scholars in phonology, phonetics, and morphology, and his vision and research have led the field to new levels of rigor and understanding.

Biography

Morris Halle was born in Liepaja, Latvia on July 23, 1923; he emigrated to the United States in 1940. He attended the City College of New York in Engineering (1941–1943), served in the United States Army in World War II (1943–1946). He received his M.A. (1948) in Linguistics from University of Chicago and did Graduate work in Slavic Languages at Columbia University (1948–1949) under the tutelage of Roman Jakobson. He did his Ph.D. (1955) on Slavic Languages, Harvard, in a dissertation entitled *the Russian Consonants: a Phonemic and Acoustical Investigation*, supervised by Roman Jakobson. He was an instructor in Russian, North Park College, Chicago, 1946–1947; Teaching Fellow in German and Russian, University of Chicago, 1947–1948; Teaching Fellow in Russian, Harvard University, 1947–1948; Assistant Professor of Modern Languages, Massachusetts Institute of Technology (MIT), 1951–1956; Associate Professor of Modern Languages, MIT, 1956–1961; Fellow, Center for Advanced Study in the Behavioral Sciences, Stanford University, 1960–1961; Founder, doctoral program in linguistics, MIT, 1961; Professor of Modern Languages, MIT, 1961–1976; Head, Department of Foreign Languages and Linguistics, MIT, 1976; Head, Department of Linguistics and Philosophy, MIT, 1976–1977; Ferrari P. Ward Professor of Modern Languages and Linguistics, MIT, 1976–1981; Institute Professor, MIT, 1981–1996; and Institute Professor Emeritus, MIT, 1996–present. He was a J.S. Guggenheim Fellow in 1960–1961 and J.R. Killian Jr. Faculty Achievement Award Lecturer at MIT in 1978–1979. He received the Science Prize from Union d'Assurances de Paris in 1991; he also received D.Sc. (hon), Brandeis University in 1988, and DHL, University of Chicago, 1992. He is a member of the

Linguistic Society of America 1951–present, and was Vice President, Linguistic Society of America in 1973; President, Linguistic Society of America in 1974; and Fellow, American Academy of Arts and Sciences 1963–present. Morris Halle is also a member of the National Academy of Sciences from 1988 to present.

References

Anderson, Stephen R, and Paul Kiparsky (eds.) 1973. A Festschrift for Morris Halle. New York: Holt, Rinehart and Winston, Inc.

Aronoff, Mark, and Richard T. Oehrle (eds.) 1984. Language sound structure: studies in phonology. Presented to Morris Halle by his teacher and students. Cambridge, MA: MIT Press.

Chomsky, Noam, and Halle Morris. 1968. The sound pattern of English. Cambridge, MA: MIT Press.

Goldsmith, John A. 1990. Autosegmental and metrical phonology. Oxford: Basil Blackwell.

Halle, Morris. 1959. The sound pattern of Russian. The Hague: Mouton Publishers.

———. 1962. Phonology in generative grammar. Word 18. 54–72.

———. 1973. Prolegomena to a theory of word formation. Linguistic Inquiry 4. 3–16.

———. 1975. Confessio Grammatici. Language 51. 525–35.

———. 1977. Tenseness, vowel shift and the phonology of back vowels in modern English. Linguistic Inquiry 8. 611–25.

———. 1983. On distinctive features and their articulatory implementation. Natural Language and Linguistic Theory 1. 95–105.

———. 1989. Why phonology is different. Linguistic Inquiry 20. 51–70.

———. 1995. Feature geometry and feature spreading. Linguistic Inquiry 26. 1–46.

———. 1997. On stress and accent in Indo-European. Language 73. 275–313.

Halle, Morris, and William Idsardi. 1994. General properties of stress and metrical structure. The handbook of phonological theory, ed. by John Goldsmith, 403–43. Oxford: Blackwell Publishers.

Halle, Morris, and Samuel J. Keyser. 1971. English stress: its form, its grows and its role in verse. New York: Harper and Row.

Halle, Morris, and Alec Marantz. 1993. Distributed morphology and the pieces of inflection. ed. by Kenneth L. Hale and Samuel J. Keyser, 11–176. The view from building 20: essays in linguistics in honor of Sylvain Bromberger.

Halle, Morris, Bert Vaux, and Andrew Wolfe. 2000. On feature spreading and the representation of place of articulation. Linguistic Inquiry 32. 387–444.

Halle, Morris, and Jean-Roger Vergnaud. 1987. An essay on stress. Cambridge, MA: MIT Press.

Harley, Heidi, and Rolf Noyer. 1999. State-of-the-article: distributed morphology. Glot International 4.4. 3–9.

Jakobson, Roman, C.G.M. Fant, and Morris Halle. 1952. Preliminaries to speech analysis. Cambridge, MA: MIT Press.

Jakobson, Roman, and Morris Halle. 1956. Fundamentals of language. The Hague: Mouton Publishers.

Kenstowicz, Michael. 1994. Phonology in generative grammar. Oxford: Blackwell Publishers.

Kenstowicz, Michael, and Charles Kisseberth. 1979. Generative phonology: description and theory. San Diego: Academic Press.

Kiparsky, Paul. 1982. From cyclic phonology to lexical phonology. The structure of phonological representations, Vol. 1. ed. by Harry van der Hurst and Norval Smith, 131–75. Dordrecht: Foris.

Mohanan, K.P. 1982. Lexical Phonology, Ph.D. Dissertation MIT, Cambridge, MA.

Newmeyer, Frederick. 1986. Linguistic theory in America. San Diego: Academic Press.

ANDREW CARNIE

See also **Chomsky, Noam; Jakobson, Roman; Morphology; Phonology**

Halliday, Michael Alexander Kirkwood

In his thesis ('The secret history of the Mongols') and in other research, M.A.K. Halliday mainly drew on what he had learned in China, but for his linguistic theory he drew on the work of the British teacher J.R. Firth, who had drawn from his in-depth knowledge of Indian Linguistic tradition. Firth was also influenced by the work of Polish anthropologist Bronislaw Malinowski. All these strands inform Halliday's work, which is unique in its multinational range of influence.

Halliday's work provides a theoretical base, offers descriptive illustration, and projects applied values in almost all branches of linguistics, grammar and semantics, discourse analysis and stylistics, phonology, sociolinguistics, computational linguistics, language education, and child language development.

The most distinguishing feature of Halliday's work is its *holistic character.* Halliday interprets language as a network of relationships and locates it in a sociocultural context. For him, 'language' is a resource for making meaning—a *semeogenic system.* He suggests that language is a higher-order semiotic system—one that has a lexicogrammatical level of organization (i.e. a system of wording) and one that is metafunctionally organized (i.e. organized functionally into

simultaneous strands of meaning: ideational, interpersonal, and textual). Our experience gets transformed into meaning and this is effected by grammar. Hence, Halliday says that the grammar of every natural language is a theory of human experience: 'a theory that we hold unconsciously, but that is all the more potent for that very reason.' Grammar 'construes our experience' but also 'enacts our interpersonal relationships'—sharing experiences with others, giving orders, making offers, and so on. In other words, 'grammar transforms experience into meaning' and it is this that constitutes what is called 'understanding'. 'Understanding and knowing are semiotic processes—processes of the development of meaning in the brain of every individual; and the powerhouse for such processes is the grammar.' In his special lecture on July 19, 1999 at the Central Institute of English and Foreign Languages in Hyderabad (India) — 'Language and the Reshaping of Human Experience'—Halliday said that

> Children of around four to six years old are just reaching the stage, in their language development, when they can handle meanings that are abstract: they can construe entities that have no perceptual correlate, like *worth* and *clue* and *habit* and *intend* and *price*; and this has two important consequences. First, it means that they can cope with abstract symbols, like letters or characters, and the abstract concepts that go with them (including, the critical distinction between writing and drawing); so they can now master this new medium. Secondly, it means that they can cope with abstract categories, and so are ready to explore new forms of knowledge. In other words they are ready for a reshaping of their previous experience.

Halliday's work on the internal organization of Language as a Semiotic system has also been holistic. He describes semantics, lexicogrammar, and phonology as resources for the language user, representing them as networks of options in what the user can mean. In his work 'system', comprising choices on the paradigmatic dimension, is the central organizing concept.

Systemic Functional theory has been applied with a wide range of contexts of research application. Halliday has pioneered the development of 'educational linguistics', interacting with teachers since the 1960s in the context of various projects. He has made significant contributions in other areas, including literary studies and computational linguistics. He was involved in work on machine translation already in the 1950s.

His work has been particularly influential in the development of text generation as one aspect of natural language processing—particularly through the Penman system that was launched in 1980 at the Information Sciences Institute of the University of Southern California. In the early 1990s, Halliday began to interact with researchers developing fuzzy theory and logic in Japan.

Halliday's commitment to research applications is related to his notion that linguists are socially accountable. This is also reflected in his concern, since the 1950s, with giving value to all varieties of language—showing, for example, that spoken language is highly systematic and complex in spite of the prevailing myth that it is scrappy and unorganized. The Hallidayan mode of intellection recommends developing multiple views on a phenomenon and advises engendering synthesis when faced with a choice between a thesis and its antithesis. He wants us to move in helixes, not in circles.

Biography

Michael Alexander Kirkwood Halliday was born in Leeds, Yorkshire (England) in 1925. He received his Ph.D. in Chinese Linguistics from Cambridge University in 1955.

He was earlier a student in China at Peking University and Lingnan University, where he assisted Professor Wang Li in surveying the dialects of Chinese in the Pearl River Delta. He taught at Edinburgh University, University College London, Nairobi (Kenya), Chicago, Stanford, and Essex University. In 1976, he joined the University of Sydney as the founding Professor of Linguistics. He retired in 1987 and went on to serve as Visiting Professor at NUS, Singapore and at ICU Tokyo.

Halliday has received honorary doctorates from the University of Nancy, the University of Birmingham, York University, and Macquarie University and was conferred the Honorary Fellowship of the Central Institute of English and Foreign Languages (Hyderabad, Lucknow, and Shillong), India. He has published over 150 books and articles. His seminal works are collected and edited by Jonathan J. Webster and Published by Continuum (London and New York), as a ten-volume series.

References

Halliday, Michael A.K. 1957. Some aspects of systematic description in grammatical analysis. Studies in linguistic analysis, 54–67. London: Basil Blackwell.

_____. 1961. Categories of the theory of grammar. Word 17. 241–92.

_____. 1967. Intonation and grammar in British English. The Hague: Mouton.

_____. 1967/1968. Notes on transitivity and theme in English. Journal of Linguistics 3. 71–81, 199–244; 4. 171–216.

_____. 1975/1977. Learning how to mean: explorations in the development of language. London: Edward Arnold; New York: American Elsevier.

_____. 1976. System and function in language, ed. by G.R. Kress. London: Oxford University Press.

_____. 1978. Language as a social semiotic: the social interpretation of language and meaning. London: Edward Arnold.

_____. 1985. An introduction to functional grammar. London: Arnold; New York: Oxford University Press (second revised edition 1994).

Halliday, Michael A.K., and Christian M.I.M. Matthiessen. 1999. Construing experience through meaning. London and New York: Continuum.

Halliday, Michael A.K., and Ruqaiya Hasan. 1976. Cohesion in English. London and New York: Longman.

<div align="right">Vᴇɴɴᴇʟᴀᴋᴀɴᴛɪ Pʀᴀᴋᴀꜱᴀᴍ</div>

See also **Firth, John Rupert**

Handwriting

Despite the common use of electronic communication with keyboards, handwriting is still a major mode of written language production. The study of handwriting is concerned with the various mechanisms and processes that intervene between the intention to write a message and the movements of the body that execute the intention by instrument gestures on a writing surface. Some of the processes, such as conceptualization and formulation, are the same for spoken and written utterances. This paper will therefore concentrate on the specific movement-control aspects of handwriting: its planning, preparation, and production. Furthermore, it will be restricted to cursive Latin script by skilled right-handed adults. When fully developed, the skill enables fluent, semiautomatic writing gestures, permitting cognitive and linguistic processing to continue in parallel.

The aim of handwriting research is to arrive at a comprehensive insight into how all aspects of skill and task are acquired, represented, planned, and organized by the writer, seen as a human information-processing system, and how the eventual kinematic, or movement, sequences are physically performed. In particular, there is interest in mutual interactions, i.e. in how the act of handwriting influences linguistic production on the one hand, and in how kinematic performance measures reveal cognitive processes on the other. The latter aspect has received the most attention in recent investigations in the multidisciplinary field of quantitative and formal handwriting research, which, in contrast to the art of graphology, is known as 'graphonomics'. Contemporary research is primarily experimental and is executed in laboratories equipped with timing devices and two- and three-dimensional movement-recording instruments.

Exact timing allows an estimate of the delay (*reaction time*) between the instruction to write and the actual start of the movement; this duration reflects the intervening amount of cognitive processing, because every cognitive process takes a measurable amount of time. Under certain assumptions, it is also possible to determine the nature of the intervening processes, e.g. letter-shape selection or force implementation. Likewise, the duration of the movement itself (*movement time*) may be assessed; this is indicative of the current *mental load* on the limited-capacity information-processing system, because of the preparation and motor programming of further parts of the message.

Because of the large forces of the involved muscles relative to the little writing movements and the small implied forces, such as friction and gravitation, it seems justified to consider the *kinematics* (i.e. the movement features) in the two-dimensional writing plane as a function of time as an adequate description of the dynamics of the handwriting trajectory. The standard two-dimensional method is to let the participant in the experiment write on a graphics tablet (*digitizer*) that samples the *x* and *y* coordinates of the pen tip at a high rate (100 times per second or more). Dedicated software analyzes the writing trace, yielding relevant parameters such as size, slant, curvature, velocity, acceleration, jerk, pen pressure, and pen tilt as a function of time. Apart from reaction time and movement time, fluency measures may also reveal local or global processing load. Many laboratories now have a three-dimensional *motion-tracking* apparatus as well, to monitor the individual joints (from shoulder to fingertip) involved in writing movements. By geometric transformation, the recorded data may be represented, e.g. as a time function of the joints' *rotation*, or as the *phase relations* between several moving effector segments (e.g. hand and fingers).

The writing limb, with its multitude of segments, joints, and muscles in the arm, hand, and fingers, is capable of moving in many different ways. Technically speaking, it has a large number of *degrees of freedom*. However, this number is often reduced—although flexibly and not always in the same way—by

the coactivation of several groups of muscles acting as single functional *synergies* or *coordinative structures*. In very general terms, the shoulder and elbow take care of the global rightward progression across the page, while the hand rotating in the wrist is responsible for the small horizontal movements within and between letters, and the thumb and fingers are mainly in charge of the vertical letter strokes.

Since 1965, mechanical and computational models of handwriting have been developed. The earliest ones were concerned with the biomechanical formation of strokes and their concatenation into letters and words. Later models, since 1980, accounted for the cerebral and neural organization of handwriting and assumed information-processing stages in the cognitive-psychology tradition. Examples of these stages are the selection and ordering of the orthographically required letters (graphemes), the retrieval of a specific letter shape (allographs; see below) from long-term motor memory, and the setting of the required muscles and their force levels to execute that letter by a selected effector (e.g. hand or fingers) in the current task-space context at a specified size and speed. Recently, the models stemming from the biomechanical tradition have gained in complexity and scope. They incorporate the control of handwriting gestures at higher levels in the brain, so that, e.g. the selection of the most suitable effector combination for smooth performance is accounted for. Thus, the two types of model are increasingly covering common ground.

From a psycholinguistic viewpoint, however, the neurocognitive and *information-processing* models have made the most interesting contributions by generating testable hypotheses on the handwriting process. These models assume that handwriting involves a hierarchical procedure in which to-be-written larger units (representing the entire message, whole phrases, and words) are activated first, followed step-by-step by the activation of their even more detailed components (single allographs and individual strokes) at increasingly lower levels. *Allographs* are the (often idiosyncratic) shape variations of a grapheme with a certain letter identity (e.g. a specific cursive lowercase <r>). Different allographs are required or preferred in different contexts. Allograph representations are assumed to be stored as generalized *motor programs* in long-term memory, and—when required in the message—ready to be retrieved and executed in any size and orientation (e.g. on a large vertical blackboard).

Strokes—effectively near-vertical trace segments—are defined in the spatial domain by the upper and lower turning points of the pen-tip excursions. These turning points are characterized by sharp bends (high *curvature*), which are associated with low speed (*velocity minima*). In fluent writing, a stroke is pro-duced *ballistically*, i.e. in a single burst, with only one acceleration and one velocity peak. The normal duration of such a fluent stroke is 100 to 125 milliseconds. A continued series of loops (each considered as two strokes, one down and one up) will thus result in a repetitive digitized signal with a *frequency* of 4 to 5 hertz. An allograph may contain between two and six strokes, so that on average, two or three letters are written per second. Allographs and strokes have both been proposed as the *basic units* of handwriting processing (long-term storage, planning, and execution); more recently, it has become clear that the size of the basic unit depends on familiarity, i.e. on learning and practice.

A representative information-processing model combines *serial* with *parallel* processing. Handwriting is seen as a multicomponent task involving various independent *modules*. From high to low, these deal with semantics (meaning), syntax (sentence structure), words, speech sounds, graphemes, allographs, strokes, task space, joints, and muscles. The order given here reflects the serial processing sequence: each module transmits its processing result to the next-lower level. The parallel features of the model originate from the principle that concurrently with the lower-order written performance of an earlier part of the message, later parts will be planned and prepared by modules higher in the system. Each module requires a small amount of *processing time* (milliseconds) and *capacity* drawn from a common source. If a message contains a processing difficulty for one of these modules, this may be reflected by a delayed initiation (increased reaction time) or by a slower ongoing writing movement (increased movement time). Because of the processing priority of larger units, increased *cognitive load* at higher levels is reflected earlier in the writing trajectory than an extra load at lower levels. An unusual allograph combination will thus be 'announced' by a slight delay several letters in advance, whereas a difficult stroke will result in a slower production of the immediately preceding stroke or even of the very stroke itself.

Research over three decades has shown that linguistic factors of various kinds and levels do indeed influence the handwriting process in a modular way, as indicated here. Initiation and trajectory formation are affected by the features of units with high, medium, or low linguistic status, which are assumedly processed at corresponding levels in the proposed handwriting-movement hierarchy. As indicated, *high* levels involve words, morphemes (subword units of meaning), and syllables; *medium* levels encompass phonology (speech-sound features), orthography (spelling), and letter clusters; *low* levels imply graphemes, allographs, and strokes. In many cases, these findings are

corroborated by clinical observations involving *double dissociation* between writing disturbances: different patients seem to suffer from different exclusive writing failures, presumably mediated by specific neural structures at different levels. Application of the current insights into the neurocognitive and motor-control aspects of handwriting has resulted in advances in educational and remedial practice, in medication for motor disorders, in written-document examination, in the analysis of ancient writing, in forensic expertise, and in automatic handwriting recognition.

References

Askov, E., W. Otto, and W. Askov. 1970. A decade of research in handwriting: progress and prospect. Journal of Educational Research 69. 86–98.

Bouwhuisen, Chris F., Ruud G.J. Meulenbroek, and Arnold J.W.M. Thomassen. 2002. A 3D motion tracking method in graphonomic research: possible applications in future handwriting recognition studies. Pattern Recognition 35. 1039–47.

Bullock, Daniel, Stephen Grossberg, and Christian Mannes. 1993. A neural network model for cursive script production. Biological Cybernetics 70. 15–28.

Ellis, Andrew W. 1993. Reading, writing, and dyslexia, 2nd edition. Hillsdale, NJ: Erlbaum.

Faure, Claudie, Paul Keuss, Guy Lorette, and Annie Vinter (eds.) 1994. Advances in handwriting and drawing: a multidisciplinary approach. Paris: Europia.

Graham, S., and N. Weintraub. 1996. A review of handwriting research: progress and prospects from 1980 to 1994. Educational Psychology Review 8. 7–87.

Kao, Henry R. S., Gerard P. van Galen, and R. Hoosain (eds.) 1986. Graphonomics: contemporary research in handwriting. Amsterdam: North-Holland.

Levelt, Willem J.M. 1989. Speaking: from intention to articulation. Cambridge, MA: MIT Press.

Martlew, Margaret (ed.) 1983. The psychology of written language. New York: Wiley.

Peck, M., E.N. Askov, and S.H. Fairchild. 1980. Another decade of research in handwriting: progress and prospects in the 1970s. Journal of Educational Research 73. 283–98.

Plamondon, Rejean, Ching Y. Suen, and Marvin L. Simner (eds.) 1989. Computer recognition and human production of handwriting. Singapore: World Scientific.

Thomassen, Arnold J.W.M., Paul J.G. Keuss, and Gerard P. van Galen (eds.) 1984. Motor aspects of handwriting: approaches to movement in graphic behavior. Amsterdam: North-Holland.

Van Galen, Gerard P. 1991. Handwriting: issues for a psychomotor theory. Human Movement Science 10. 165–91.

ARNOLD J.W.M. THOMASSEN

Harris, Zellig Sabbetai

The scientific legacy of Zellig S. Harris can be divided into five areas: (1) Northwest Semitic linguistics; (2) Structural analysis; (3) Transformational analysis; (4) Operator grammar; and (5) Politics.

While Harris was an undergraduate student at the University of Pennsylvania in 1929, epoch-making texts in Ugaritic were published and promptly deciphered. This turned his attention to new and exciting possibilities of research in Northwest Semitic, and he subsequently worked in this area with his teacher J.A. Montgomery, publishing with him a work on Ugaritic mythological texts. He already had a background in Hebrew from his family. During the 1930s, he established a solid reputation as a semitist, publishing a Phoenician grammar in 1936, and a study of Canaanite dialects in 1939. Both works have remained useful. It was here that he developed the style of writing that lasted throughout his career. It was direct and sparse, with a wealth of detail and documentation when these were called for. But he made no concessions to the reader. Harris did not pretend to be a popularizer, and one may have to struggle to follow his technical, tightly constructed argumentation.

He then moved into a more general study of language theory, even though he continued to maintain an interest in Semitic languages as material for theoretical investigations. Other semitists regretted his defection from the field, and even expressed resentment at what he was doing. The Semitics scholar J. Cantineau, in a review of Harris's book *Methods in structural linguistics,* claimed that Harris placed facts on a procrustean bed to fit his theories. He added that Harris brought nothing new or useful to linguistics, despite the specious appearance of being scientific.

Harris shook off all criticism and entered his new phase of scholarship, the cornerstone of which was the book that Cantineau reviewed unfavorably. Although he was not formally a student of Edward Sapir, he acknowledges his great debt to him. Harris stresses the importance of describing language in terms of how items fit together, and how they are distributed. While he admits that meaning is the goal of most language (poetry may sometimes be an exception), it is not susceptible to scientific quantification or description, and it, along with psychological and mentalistic notions, must be strictly avoided. He tried to isolate

the elements of a language, and the relationships that were possible between them. Harris adhered to the structuralists' credo that linguistics best studies language synchronically, and on its own terms.

Up to his time, analysis of language, whatever its theoretical basis might be, ended at the conclusion of the sentence. Harris had the ambition of setting up an analysis of texts, sentences that were strung together, and this led to his concept of Discourse Analysis, a subject on which he taught a course regularly at the university. This endeavor led him to set up normal forms of sentences, so that they could be compared within texts, and this brought him, more or less accidentally, to the notion of transformations that would connect, for example, a sentence in its active and passive form (the dog bit the man/the man was bitten by the dog). Thus evolved the idea of 'transformations', which led in turn to 'transformational grammar', which has acquired great prominence, mainly on account of the research and publications of Noam Chomsky, who was Harris's student. This gave rise to a school of thought of continuing prominence, which was subjected to obloquy similar to that meted out once to the structuralists, being described by the structuralist Charles F. Hockett (died: 2000) as 'a theory spawned by a generation of vipers'.

But Harris did not go down the same path as Chomsky. Already in the 1960s, he was teaching a course in Linguistics and Mathematical Logic, in which he was mulling over his aim to produce a mathematical model for language. It should be stressed that Harris, in moving to his new approach to language, which was distinct from the path he had previously followed, and distinct from that of Chomsky, did not thereby reject his previous work, even though he may have felt that he had taken it as far as he reasonably could. Harris did not belong to a 'God's truth' mode of looking at scientific description. Various approaches might have their own particular validity. He was remarkably unpartisan, and spent no time in engaging in public relations on behalf of himself or his views, a fact that may account for the relative obscurity of his viewpoints, and his failure to establish a school devoted to his principles.

His new approach is best understood by a simple example. If we consider the sentence *sheep eat grass,* acceptable to any English speaker, Harris observes that the word *eat* imperatively requires two other words, which do not themselves require anything else. He expresses this as O_{nn}, where O represents the operator (in this case the word *eat)* and this operator requires two arguments (here *sheep* and *grass)* which can be considered to be operators with null arguments, hence *n.* If we expand this sentence to *I know that sheep eat grass,* then *eat* now has an operator as its

first argument ('I know') and the second argument is the same, giving O_{on}. Harris recognizes that this may not give an analysis of pedagogic value by way of giving a useful order of the findings; but then he was not particularly interested in pedagogy. Like Everest climbers, he explored the claimed mathematical structure of language simply because it was there. Harris also explores the unreduced sentences of a particular language. We have no evidence of the existence now of a 'me Tarzan, you Jane' type of prehistoric language; so-called primitive languages may have richly complex syntactic structures, and are not fundamentally different in nature from other better known languages. Maybe he considered these unreduced sentences to belong to an earlier stage of language, but he does not state this expressly, even though toward the end of his life in his masterly *A theory of language and information*, he gingerly explores the nature and development of language. A simple example may be seen in *sheep eat green grass.* This is a reduced form of two sentences *sheep eat grass* and *grass is green,* i.e. *sheep eat grass which is green,* where the first two letters of *which* (*wh...*) form a link, and the remainder represents the word *grass.* An element of economy plays a part here, which has long been recognized as a possible factor in phonetic change. Harris believes that the operator-argument method is fully adequate for unreduced sentences at least. *A theory of language and information,* represents the end of his life work, which had remarkable consistency and tenacity. Thereafter, he devoted himself to a substantial political tract, published posthumously by a group of his admirers.

Biography

Zellig Sabbetai Harris was born in Balta, Ukraine, on October 12, 1909. He was brought to Philadelphia at the age of four years. His father was a *mohel* (Jewish ritual circumciser) and cantor. Harris received the degree of A.B. at the University of Pennsylvania in 1930, followed by a Ph.D. four years later, and spent his entire academic career at that institution as a student and faculty member, until his retirement at the age of 70. For much of his life he commuted between Philadelphia and Israel, where he was a member of the communal settlement Kibbutz Mishmar Haemek, near Haifa, and reportedly was known there simply as Zellig the carpenter. This kibbutz was part of the extreme left-wing *Hashomer Hatzair* ('the Young Watchman') movement. He participated in setting up an adult education program for members of such settlements. His wife Bruriah was a researcher in physics at the Weizmann Institute in Rehovot, Israel. Harris kept his academic and political interests separate, and

in his scientific work never broached his political views, which were closest to anarchism in the spirit of Kropotkin. In 1941, he took over the editorship of the *Journal of the American Oriental Society*, one of the oldest learned societies in the United States, from Norman Brown, a colleague at the University of Pennsylvania. He served in that capacity until 1948. He was president of the Linguistic Society of America in 1955. He published numerous articles in the journal *Language* until 1957; after that time, most of his books and papers were published in Europe. He died on May 22, 1992, just after publishing an uncharacteristically readable account of his views in his *A theory of language and information.*

References

Chomsky, Noam. 1957. Syntactic structures. The Hague: Mouton. Janua Linguarum, Series minor, nr. 4, 117p.

Harris, Zelig Sabbetai. 1939. Development of the Canaanite dialects; an investigation in linguistic history. New Haven: American Oriental Society. American Oriental series, v. 16. x, 108p.

_____. 1951. Methods in structural linguistics. Chicago: University of Chicago Press. xv, 384p. Published in 1960 under title: Structural linguistics.

_____. 1952. Discourse analysis. Language 28. 1–30.

_____. 1957. Co-occurrence and transformation in linguistic structure. Language 33. 283–340.

_____. 1982. A grammar of English on mathematical principles. New York: Wiley, xvi, 429p.

_____. 1991. A theory of language and information: a mathematical approach, Oxford: Clarendon Press; New York: Oxford University Press. xi, 428p.

_____. 1997. The transformation of capitalist society. Lanham, Maryland: Rowman and Littlefield. [Completed manuscript prepared for publication posthumously by Murray Eden, William Evan, and Seymour Melman.]

Hiź, Henry. 1994. Zellig S. Harris. Proceedings of the American Philosophical Society 138. 509–22.

Koerner, E. F. Konrad, 1993. Zellig Sabbetai Harris: a comprehensive bibliography of his writings, 1932–1991. Historiographia Linguistica 20: 2/3 pp. 509–22. An updated version of this bibliography is on the World Wide Web at http://www.dmi.columbia.edu/zellig/ZSHbibliography.html. This includes a bibliography of evaluations of Harris's work.

Matthews, Peter. 1999. Zellig Sabbetai Harris. Language 75. 112–9.

ALAN D. CORRÉ

Haugen, Einar

Einar Haugen was one of the foremost Scandinavian linguists of his time. His parents both immigrated from central Norway to the United States, settling in Sioux City, Iowa, in 1899. Haugen grew up there, in a Norwegian neighborhood, and learned Norwegian as a native language and English in school. He in fact learned two dialects of Norwegian, the more standard dialect spoken by his schoolteacher mother and the more rural dialect spoken by his father. He spent two years in Norway as a child, from 1914 to 1916, between the ages of eight and ten, which allowed him to master his father's dialect in its natural setting. This stay reinforced his already strong ties to Norway and Norwegian; the influence of his Norwegian immigrant community and his childhood visit to Norway cannot be underestimated when analyzing his life's work.

Haugen contributed to many areas of linguistics. He was deeply concerned with the Norwegian language, writing several textbooks of Norwegian, and directing the compilation of a Norwegian–English dictionary. During World War II, he directed a project at the University of Wisconsin to teach servicemen Norwegian. He also contributed extensively to Norwegian dialectology, publishing, among other works, a two-volume survey of varieties of Norwegian spoken in the United States (*The Norwegian language in America: a study in bilingual behavior*; originally published in two volumes in 1953 and reissued as one volume in 1969) and numerous articles on the topic.

He was also interested in numerous aspects of sociolinguistics. His interest in language planning and language standardization goes back to his dissertation; he continued to work in this field, and eventually published a book dealing with language planning and Norwegian. Bilingualism, another interest stemming from his childhood experiences, was also a prominent theme in his work. *The Norwegian language in America*, as indicated by the subtitle, also dealt extensively with bilingualism, focusing largely on the adaptation of English words into American Norwegian. A later volume, *Bilingualism in the Americas: a bibliography and research guide*, was also a seminal contribution to the field. Finally, Haugen was an important contributor to the field of 'language ecology' (and in fact coined the term), the study of interactions between a language and its environment. His 1972 work,

443

The ecology of language, reprints selected essays on topics such as the relationship between language and immigration and communication difficulties between speakers of various languages.

Haugen's contributions to historical Scandinavian linguistics are critical to the field. He prepared an edition and translation of *The first grammatical treatise*, an anonymous twelfth-century Icelandic work on Old Norse phonology and orthography. His monumental work on the history of the Scandinavian languages, *The Scandinavian languages: an introduction to their history*, originally published in 1976 and translated into German in 1984, remains a standard handbook. His later monograph, *Scandinavian language structures: a comparative historical survey*, offers an accessible discussion of the development of the various components (phonology, morphology, syntax, and the lexicon) of the Scandinavian languages.

Haugen also helped shape the field of structuralist linguistics. He authored two influential articles on structuralist phonology, 'Facts and phonemics' (with W.F. Twaddell 1942) and 'Phoneme or prosodeme' (1949). Other phonological topics he dealt with included the vowel systems of Norwegian and Icelandic, tone in Norwegian, and the role of the syllable in phonology. Yet, he was never parochial in his scholarship; in his presidential address to the Linguistic Society of America (published in 1951), he urged American linguists to avoid scholarly isolationism and take developments from European scholarship into consideration.

Haugen also contributed to literary studies. He translated a number of Icelandic sagas and plays, a history of Norwegian literature, and a biography of Henrik Ibsen. He prepared word counts of a number of Scandinavian authors, and also published studies of various Icelandic sagas.

Biography

Einar Haugen was born in Sioux City, Iowa on April 19, 1906. He attended Morningside College in 1924–1927. He received his B.A. in 1928 from St. Olaf's College; M.A. in 1929, and Ph.D. for a dissertation on the development of the New-Norwegian dialect of Norwegian, University of Illinois (Urbana) in 1931. He was Assistant Professor, University of Wisconsin (Madison), 1931–1936; Associate Professor, 1936–1938; Torger Thompson Professor of Scandinavian Languages, 1938–1962; Vilas Research Professor of Scandinavian Languages, 1962–1964; Victor S. Thomas Professor of Scandinavian and Linguistics, Harvard University, 1964–1975; Cultural relations attaché, US Embassy in Oslo, 1945–1946; and President, Society for the Advancement of Scandinavian Study, 1938; Linguistic Society of America, 1950; Ninth International Congress of Linguists, 1962; American Dialect Society, 1962. He was also a member of Permanent International Committee of Linguists in 1966–1972. He was visiting lecturer (1938) and Fulbright Research Professor (1951–1952), University of Oslo; Visiting professor, University of Iceland, 1955–56; Adviser to the English Language Educational Council of Japan, 1958–1960; and Fulbright lecturer, University of Uppsala, 1976–1977. He received grants and fellowships from the Guggenheim Foundation (1942), National Endowment for the Humanities (1967), Center for Advanced Study in the Behavioral Sciences at Stanford University (1963–1964), and National Science Foundation (1967–1969). Haugen was a member of the American Academy of Arts and Sciences, the Norwegian, Icelandic, Swedish, and Danish academies of science, and the Linguistic Research Group of Pakistan. He also received the Order of St. Olaf, Norway in 1940 and Order of the North Star, Sweden in 1961. He received several honorary degrees, including University of Michigan (1953), St. Olaf College (1958), the University of Oslo (1961), and the University of Iceland (1971). He also received the Nansen Award in Oslo in 1970 and the Jancke Prize in Uppsala in 1976. Haugen died in Cambridge, Massachusetts, on June 20, 1994.

References

Appel, René, and Pieter Muysken. 1987. Language contact and bilingualism. London: Arnold.

Eliasson, Stig, and Ernst Håkon Jahr (eds.) 1997. Language and its ecology. Essays in memory of Einar Haugen. Berlin: Mouton de Gruyter.

Firchow, Evelyn Scherabon, Kaaren Grimstad, Nils Hasselmo, and Wayne O'Neil (eds.) 1972. Studies for Einar Haugen. The Hague: Mouton.

Fishman, Joshua. 1995. Einar Haugen. Language 71.

Haugen, Einar, and W.F. Twaddell. 1942. Facts and phonemics. Language 18.

Haugen, Einar. 1947. Spoken Norwegian.

———. 1949. Phoneme or prosodeme. Language 25.

———. 1950. First grammatical treatise: the earliest Germanic phonology. An edition, translation, and commentary; 2nd revised edition 1972.

———. 1951. Directions in modern linguistics. Language 27.

———. 1953. The Norwegian language in America: a study in bilingual behavior. Vol. I: The bilingual community. Vol. II: The American dialects of Norwegian; republished in one volume, 1969.

———. 1956. Bilingualism in the Americas: a bibliography and research guide.

———. 1972. The ecology of language.

———. 1972. Studies by Einar Haugen.

———. 1976. The Scandinavian languages: an introduction to their history.

———. 1982. Scandinavian language structures: a comparative historical survey.

———. 1987. Blessings of Babel : bilingualism and language planning. Problems and pleasures.

Naess, Harald S., and Sigmund Skard (eds.) 1971. Studies in Scandinavian–American interrelations: dedicated to Einar Haugen. Oslo: Universitetsforlaget.

Romaine, Suzanne. 1989. Bilingualism. Oxford: Blackwell.

MARC PIERCE

See also **Bilingualism; Language Planning; Structuralism**

Hausa and Chadic Languages

The Chadic language family includes an estimated 140–150 languages spoken in the sub-Saharan area to the west, south, and east of Lake Chad. The best known and most widely spoken Chadic language is Hausa, with upward of 30 million first-language speakers, more than any other language in Africa south of the Sahara. The remaining languages, some of which are on the brink of extinction, probably number little more than several million speakers in total, varying in size from less than half a million to just a few dozen speakers, and new languages are still occasionally discovered. Most Chadic languages remain unwritten, and for many only brief word-lists are available, although for some, e.g. Bidiya, Guruntum, Kanakuru, Kera, Kwami, Lamang, Margi, Miya, and Mupun, good descriptive grammars now exist, and several high-quality dictionaries have appeared, e.g. Dangaléat, Lamé, Ngizim, and Tangale. Hausa has four up-to-date and comprehensive reference grammars (all produced within the last few years) and two first-class dictionaries, and these works together make it the best documented language in sub-Saharan Africa.

Chadic is a member of the Afroasiatic language superfamily, which includes Semitic (e.g. Amharic, Arabic, Hebrew), Cushitic (e.g. Oromo, Somali), Omotic (e.g. Maale, Wolaytta), Berber (e.g. Tamahaq, Tamazight), and (extinct) Ancient Egyptian/Coptic. The inclusion of Chadic within Afroasiatic was first proposed almost 150 years ago, based on specific points of resemblance between individual Chadic languages and other Afroasiatic languages. These features include an *mV-* prefix (V stands for an unspecified vowel), which occurs on nouns that refer to the agent of an action, a location, or an instrument, e.g. [Hausa] *ma-hàifii* 'parent', *ma-fakaa* 'shelter', *ma-kullii* 'key, lock'. (Transcription: *aa* = long vowel, *a* = short; *à(a)* = L(ow) tone, *â(a)* = F(alling) tone; H(igh) tone is unmarked.) Despite this, the family relationship has only been generally accepted as a fact over the last 40 years or so, following Joseph Greenberg's (1963) comprehensive reclassification of African languages.

The accepted internal classification groups Chadic languages into three major coordinate branches, West, Central (= Biu-Mandara), and East, plus an isolated Masa cluster (with sub-branches and smaller groupings) as follows:

West-A, e.g. Angas, Bole, Fyer, Galambu, Gera, Goemai, Gwandara, Hausa, Kanakuru, Karekare, Kirfi, Kulere, Kwami, Mupun, Pero, Ron, Sha, Sura, Tangale

West-B, e.g. Bade, Boghom, Duwai, Geji, Guruntum, Miya, Ngizim, Pa'a, Saya, Warji

Central-A, e.g. Bachama-Bata, Bura-Pabir, Daba, Ga'anda, Gisiga, Glavda, Gude, Guduf, Higi-Kapsiki, Kilba, Lamang, Mandara, Margi, Matakam, Mofu-Gudur, Ouldeme, Sukur, Tera

Central-B, e.g. Buduma, Kotoko, Logone, Mbara, Musgu

Central-C, e.g. Gidar

East-A, e.g. Kera, Lele, Nancere, Sarwa, Somrai, Tumak

East-B, e.g. Bidiya, Dangaléat, Jegu, Mahwa, Migama, Mogum, Mubi, Mukulu, Sokoro, Toram

Masa (isolated cluster), e.g. Lamé, Marba-Monogoy, Masa, Musey

Chadic languages have complex consonant inventories. A typical pattern is for a language to distinguish voiceless stops such as /t/ from voiced /d/, in addition to implosive stops /ɗ/ and prenasalized stops /nd/.

Vowel inventories vary from two vowels, high /ə/ and low /a/, as in Bachama and Mandara, to seven vowels /i e ɛ a ɔ o u/, as in Dangaléat, with /i (e) a ə (o) u/ a regular pattern, and the diphthongs /ai/ and /au/ are found. It is also common throughout the family to distinguish between long and short vowels, especially in word-medial position.

Chadic languages have contrastive tones (distinctive intonational pitch levels), ranging from two (e.g. Hausa) or three (e.g. Angas) level tone systems. Although tone can function in lexical discrimination, its key role is normally grammatical, e.g. [Hausa] *taa/tâa/tà zoo* 'she came/will come/should come', *zàunaa!* 'sit down!' vs. *zaunàa* 'to sit down', *nân* 'here (near me the speaker)'

445

vs. *nan* 'there (near you the hearer)'. In a number of (West) Chadic languages, phonological tone is sensitive to syllable structure in certain word categories. In Bole, for example, tone on bisyllabic verbs is predictable: if the verb is C(onsonant)V(owel).CV (light first syllable) it will be H-H tone, e.g. *tonu* 'sharpen'; if it is CVC.CV or CVV.CV (heavy first syllable) it will be L-H tone, e.g. *ràamu* 'repair', *wùndu* 'call'.

Many Chadic languages have masculine/feminine grammatical gender (an inherited Afroasiatic feature), with no distinction in the plural, and typically distinguish gender in second- and third-person singular pronouns, e.g. (independent pronouns) *kai/kee* 'you (masc./fem.)', *shii/ita* 'he/she' [Hausa], *fiy/macə* 'you (masc./fem.)', *tə/njə* 'he/she' [Miya]. Overt gender marking on nouns has also evolved in a number of West Chadic languages, e.g. [Western Bade] *mângaa-n* 'my friend (masc.)' vs. *mânga-n* 'my friend (fem.)' (*-aa* = masculine, *-a* = feminine), [Hausa] *jàakii* 'donkey' vs. *jàakaa* 'she ass' (*-ii* = masculine, *-aa* = feminine). Some also exhibit the characteristic Afroasiatic *n/t/n* (masc./fem./pl.) noun/agreement marking pattern (where the masculine and plural markers are phonologically identical), cf. [Masa] *vèt-na* 'rabbit', *vèt-ta* 'female rabbit', *vèdai-na* 'rabbits'.

Noun plural formations are varied and elaborate, and some plural suffixes can be reconstructed for the presumed ancestor language Proto-Chadic, e.g. *-Vn, *-aki, *-i,* and *-ai* (with *-Vn* and *-ai* probably even reconstructable for Proto-Afroasiatic, the superfamily ancestor language). Examples: (*-Vn*) *kùmən/kùmənən* 'mouse/mice' [Bade], *miyò/mishan* 'co-wife/co-wives' [Kanakuru], (*-aki*) *goonaa/gòonàkii* 'farm(s)' [Hausa], (*-i*) *duwimà/dùwìmi* 'guineafowl(s)' [Gera], (*-ai*) *mùtù/mutai* 'sore(s)' [Dangaléat].

Verbs in many Chadic languages preserve the lexically arbitrary Proto-Chadic distinction between final *-a* and final *-ə* verbs (where the final schwa vowel is often pronounced as [i], [ə], or [u]), cf. [Tera] *na* 'see' and *dlə* 'get', [Guruntum] *daa* 'sit' and *shi* 'eat', [Bole] *pata* 'go out' and *poru* 'say', and in many West Chadic languages final *-a(a)* stems are typically limited to bisyllabic verbs with an initial light syllable, e.g. [Bole] *kù.maa* 'hear', [Hausa] *nù.ka* 'ripen after storing'. Verbal meanings are modified by the addition of one or more suffixes. These extensions encode notions such as action 'in the direction of' (centripetal) or 'away from' (centrifugal), or action 'partially completed' or 'totally completed', e.g. (centripetal) *fit-oo* 'come out' from *fìta* 'go out' [Hausa], (totality) *sà-nyà* 'drink up' from *sà* 'drink' [Margi]. Singular imperatives are often overtly signaled with an *-i* or *-u* suffix, e.g. [Bachama] *gur-ì!* 'sing!', [Ga'anda] *còk-u!* 'break (it)!'.

In some cases, the verbal suffixes also have a syntactic function, signaling, *inter alia,* transitivization (allowing a verb to take an object), or perfectivity (indicating completed action), e.g. (transitivization) *yàw-tu* 'take down' from *yàwwu* 'go down' [Bole], (perfectivity) *pet-inà* 'he came out' [Pero]. Another characteristic feature is the formation of so-called 'pluractional' verbs, denoting an action repeated many times or affecting a plurality of subjects (if intransitive) or objects (if transitive), e.g. [Hausa] *tunàa/tun-tùnaa* 'remind/remind many people and/or often', [Bole] *dàppu/dàdàppu* 'repair', [Angas] *fwin/fwan* 'untie', [Pero] *lofò/loffò* 'beat'.

In a number of languages, intransitive verbs (verbs without objects) are followed by an 'intransitive copy pronoun', which copies the person, number, and gender features of the subject, e.g. [Kanakuru] *nà pòrò-no* 'I went out' (lit. I went out-I). Verb stems can also be overtly inflected for tense–aspect–mood by segmental and/or tone changes, e.g. (perfective/subjunctive/imperfective) *kəma/kəmì/kəmà* 'hear' [Ngizim], and grammatically-required number agreement with the subject is also found sporadically, e.g. [Karekare] *mù ngàt-ʔan kò* 'we fell' (with *-ʔan* plural suffix), cf. *nà ngàtaa kò* 'I fell'. In some Chadic languages so-called 'logophoric' pronouns are used to denote the person whose utterance is represented in reported speech, e.g. (tones not marked) *wur sat ni ɗi n nas an* 'he_i said he_i will beat me' [Mupun], where the 3rd masculine singular logophoric pronoun *ɗi* signals that the subject of the subordinate clause refers to the same person as the main clause subject. Some languages also use impersonal forms (equivalent to implied [null] subjects) with arbitrary, often human, reference, to express impersonal subjects, e.g. [Miya] *a sənà-tlən* 'one will spend the night', and impersonal constructions containing transitive verbs typically correspond to agentless passives in English, e.g. [Hausa] *an dafà àbinci* 'the food has been cooked' (lit. (some)one.perf. cook food).

Word order is normally S(ubject)-V(erb)-O(bject), e.g. [Hausa] *Audù yaa sàyi mootàa* 'Audu bought a car' (lit. Audu he.perf. buy car), although VSO order is found in a few Central Chadic/Biu-Mandara languages spoken in the Nigeria-Cameroon border area, e.g. (tones not marked) *kə bilə Musa tə bwaja* 'Musa has killed the leopard' (lit. perf. kill Musa leopard) [Gude], *daxtaatə́-l marakw* 'he/one will take the woman' (lit. take.fut.-he woman) [Lamang]. Pronominal indirect objects usually attach to the verb like an affix, whereas nominal indirect objects occur as prepositional phrases to the right of the direct object, cf. [Kanakuru] *à jòɓ-rò landài* 'he washed the robe for her' (lit. he washed-for her robe), and *à jòɓè landài gən tamno* 'he washed the robe for the woman' (lit. he washed robe for woman). (Hausa is atypical in placing both pronominal *and* nominal indirect objects immediately after the verb.)

Wh-questions, focus, and relativization often pattern together with special (focus) marking. In Hausa, the *wh*-expression, focused element, or relativized head

occur at the left periphery of the clause and focus marking appears on the preverbal inflectional element, e.g. (*wh*-fronting) *wàa kukà ganii?* 'whom did you (pl.) see?' (lit. who 2pl.focus.perf. see), (focus-fronting) *yaarònkà mukà ganii* ('it's) your boy (that) we saw' (lit. boy.your 1pl.focus.perf. see), (relativization with fronting) *yaaròn dà mukà ganii* 'the boy whom we saw' (lit. boy.the rel. 1pl.focus.perf. see). In Miya questioned and focused (preverbal) subjects remain *in situ* but require special inflectional marking, e.g. *wàa də bəsàwà?* 'who washed (it)?' (lit. who foc. washed), *tə̀ də bəsàw* '*he* washed (it)' (lit.he foc. washed).

Negation in Chadic is typically signaled with a single marker in sentence-final position, e.g. [Guruntum] *tâa kyur shau dà* 'she will not cook the food' (lit. she will cook food neg.), [Kera] *wə gùsnə̀ hàrga bà* 'he didn't buy her a goat' (lit. he bought her goat neg.), sometimes accompanied by an additional reinforcing preverbal negative marker, e.g. [Hausa] *bà tà zoo ba* 'she did not come' (lit. neg. she.perf come neg.), [Margi] *nì ndà wì mài* 'I did not run' (lit. I neg. run. perf. neg.), [Miya] *fà tà bəta zhaak-uw* 'don't untie the donkey' (lit. you (masc.) neg. untie donkey-neg.).

Cognate complement constructions are also widespread in Chadic, where the verb is repeated in a morphologically related (verbo-nominal) form, e.g. [Hausa] *taa màaree shì maarìi mài ciiwòo* 'she slapped him painfully' (lit. she.perf. slap him slapping with pain), [Miya] *mən rà-ya ràw* 'I surpass him' (lit. I surpass-him surpassing). Comparatives are normally verbal sentences using the lexical verb 'exceed, surpass, be more than', e.g. [Hausa] *Audù yaa fi Muusaa wàayoo* 'Audu is more clever than Musa' (lit. Audu he.perf. surpass Musa cleverness), [Miya] *mə̀n jiy baa də̀ rà-tlən aa wàshàshàm* 'it was I who was older than them' (lit. I it is one who perf. exceed-them for year), [Margi] *nàjà gà mdia-ɗà də dzə̀gàm* 'he is taller than me' (lit. he surpassed-me with tall).

In noun phrase syntax, the normative order for constituents is head-initial, i.e. head noun followed by definite determiners, possessives, numerals, relative clauses, adjectives, etc. The linear order in genitive constructions is possessee X (+ 'of') + possessor Y, e.g. [Hausa] *yaarò-n Muusaa* 'Musa's boy' (lit. boy-of Musa), [Margi] *tagu gə Haman* 'Haman's horse' (lit. horse of Haman). Many Chadic languages also make an overt distinction between alienable and inalienable possession, whereby inalienable possession (items that belong to someone in a deeper sense—they cannot be taken away) is expressed by direct juxtaposition (i.e. with no overt linker), cf. (inalienable) *mənda Miyim* 'Miyim's wife' (lit. wife Miyim), and alienable (items that can be taken away) *gam ma tamnoi* 'the woman's ram' (lit. ram of woman) [Kanakuru].

Some languages restrict plural marking to a narrow range of nouns (typically humans and animals), and a few languages signal plurality only by modifying the invariant noun with a demonstrative, quantifier or numeral, cf. [Pero] *landàa mijè* 'that shirt' (lit. shirt that) and *landàa mè* 'those shirts' (lit. shirt those). Reflexive pronouns ('...self') and reciprocals ('each other') are typically formed with the body-part nouns 'head' and 'body' respectively, e.g. [Hausa] *kâ-n-tà* 'herself' (lit. head-of-her), [Kwami] *kuu-nì* 'himself' (lit. head-his), [Hausa] *juuna-n-mù* 'each other (we)' (lit. body-of-our), [Miya] *tuwatùwà-tlə̀n* 'each other (they)' (lit. body-their).

References

Alio, Khalil. 1986. Essai de description de la langue bidiya du Guéra (Tchad). Phonologie—Grammaire. Marburger Studien zur Afrika- und Asienkunde, Serie A, Afrika, 45. Berlin: Dietrich Reimer.

Caron, Bernard. 1991. Le haoussa de l'Ader. Sprache und Oralität in Afrika, Vol. 10. Berlin: Dietrich Reimer.

Frajzyngier, Zygmunt. 1993. A grammar of Mupun. Sprache und Oralität in Afrika, Vol. 14. Berlin: Dietrich Reimer.

Greenberg, Joseph H. 1963. The languages of Africa. (Indiana University Research Center in Anthropology, Folklore, and Linguistics, publication 25; International Journal of American Linguistics, 29:1, part 2.) Bloomington: Indiana University.

Hoffmann, Carl. 1963. A grammar of the Margi language. London: Oxford University Press.

Jaggar, Philip J. 2001. Hausa. London Oriental and African Language Library, Vol. 7. Amsterdam and Philadelphia: Benjamins.

Jaggar, Philip J., and H. Ekkehard Wolff. 2002. Chadic and Hausa linguistics: the selected papers of Paul Newman with commentaries. Cologne: Rüdiger Köppe.

Jungraithmayr, Herrmann, in collaboration with Njeno Andirya Galadima and Ulrich Kleinewillinghöfer. 1991. A dictionary of the Tangale language (Kaltungo, Northern Nigeria). Sprache und Oralität in Afrika, Vol. 12. Berlin: Dietrich Reimer.

Newman, Paul. 1980. The classification of Chadic within Afroasiatic. Leiden: Universitaire Pers.

Newman, Paul. 2000. The Hausa language: an encyclopedic reference grammar. New Haven and London: Yale University Press.

Newman, Paul. 2003. Chadic languages. International Encyclopedia of Linguistics, 2nd ed., Vol. 1. 304–7. Oxford and New York: Oxford University Press.

Pawlak, Nina. 1994. Syntactic markers in Chadic: a study on development of grammatical morphemes [sic]. Warsaw: Instytut Orientalistyczny, Uniwersytetu Warszawskiego.

Schuh, Russell G. 1998. A grammar of Miya. (University of California Publications in Linguistics, 130.) Berkeley and Los Angeles: University of California Press.

Stolbova, Olga V. 1996. Studies in Chadic comparative phonology. Moscow: Diaphragma Publishers.

Wolff, H. Ekkehard. 1983. A grammar of the Lamang language (Gwa'd Lamang). Afrikanistische Forschungen, 10. Glückstadt: J. J. Augustin.

Wolff, H. Ekkehard. 1993. Referenzgrammatik des Hausa. (Hamburger Beiträge zur Afrikanistik, 2.) Münster and Hamburg: LIT.

PHILIP J. JAGGAR

See also **Afroasiatic; Greenberg, Joseph Harold**

Hawaiian Creole English

Hawaiian Creole English (referred to by linguists as Hawaii Creole English) is one of a number of English-lexifier pidgins and creoles. Varieties of pidgin and creole English in Hawaii are the outcome of contact between Hawaiians, Europeans (primarily English speakers, who contributed most of the vocabulary to the emerging pidgin), and the various immigrant groups (e.g. Chinese, Japanese, Portuguese, and Filipinos) brought to Hawaii to work as indentured laborers on plantations from the 1850s onward. Hawaii Pidgin English, formed between 1890 and 1910, is spoken only by a few remaining members of the oldest generation of immigrants to the plantations and is now dying out. Its descendant, Hawaii Creole English, a more complex language, emerged between 1910 and 1930. Through Hawaii Creole English, the culture, identity, and solidarity of the local working class were created. It became a badge separating locals from the tourists as well as from the resident middle class English-speaking *haoles* (Hawaiian 'foreigner'), a term referring to whites.

Varieties of Hawaii Pidgin and Creole English are locally called simply 'Pidgin', even though most of them are technically forms of creole English because they function as the native rather than the second language of most of their speakers. Hawaii Creole English is the first language of the majority of locally born children and the first language of somewhat less than half the state of Hawaii's population of just over a million. The present ethnic composition of the Hawaiian islands reflects the diversity of the different waves of foreign labor imported to work on sugar plantations. Hawaii is the only US state in which an Asian/Pacific population is the majority: 61.9% of a total population of 1.1 million. Currently, those of white ancestry account for 33.4%, Japanese for 22.3%, Filipino for 15.2%, Chinese for 6.2%, and others for 10.2%. Native Hawaiians make up another 12.5% of the population.

Like most of the world's pidgins and creoles, Hawaii Creole English has no official recognition, and its coexistence with standard English has not been peaceful. Many speakers were and still are corrected at school for speaking 'bad' or 'broken' English. Standard English was seen as a tool of assimilation to a common American culture, even before pidgin formed and spread. As the emerging middle class increasingly adopted and identified with Standard English, at least within the context of the school, the use of Hawaii Creole English began to be perceived as a liability on the job market because it indicated low social status. It was a reminder of the plantation origins that many wanted to leave behind. The campaign to eradicate Hawaii Creole English goes back at least to a federal report issued in 1920 on education in Hawaii (then a US territory). The report recommended a number of reforms, among them the 'elimination of the jargon of the plantations and 'pidgin English' of the streets'. Creole speakers were discriminated against through education in a school system that was originally set up to keep out those who could not pass an English test. In this way, it was hoped to restrict the admission of nonwhite children into the so-called English Standard schools set up in 1924 and attended mainly by white children. By institutionalizing what was essentially racial discrimination along linguistic lines, the schools managed to keep creole speakers 'in their place' and maintain a distance between them and English speakers until after World War II.

As a result of the increasing contact between creole speakers and speakers of mainland varieties of English after World War II and the political incorporation of the islands into the United States as the 50th state in 1959, the boundaries between standard English and the creole have become fuzzier than elsewhere in the Pacific, e.g. Papua New Guinea. Contact with English has decreased the autonomy of Hawaii Creole English and has created a continuum of varieties resulting from changes introduced through decreolization, a process that has made the language less creole-like and more like English in some, but not all, respects. Although adjacent varieties of the continuum are mutually intelligible, the two extreme endpoints may often not be.

ai go liv om	most creole-like
autsaid fo yu	
ai gon liv om	
autsaid fo yu	
ai going liv om	
autsaid fo yu	
aim gonna liv om	
autsaid fo yu	
ail liv om/it	least creole-like
autsaid fo yu	
I will leave it/them	most standard
outside for you	English-like

The most decreolized varieties are found on the island of Oahu, where three fourths of the state's

population is located, along with the capital, Honolulu, and the main US military base, Pearl Harbor. The outer islands of Kauai and Hawaii are the least decreolized.

Although many opponents of Hawaii Creole English believed and hoped that the language would gradually be absorbed into English through continuing decreolization, this is clearly not happening. On the contrary, Hawaii Creole English is being maintained, and even strengthened, particularly among certain groups of young people, as a symbol of local identity in the face of inundation from mainland norms. In the face of public opposition to a proposed policy of allowing teachers and students to speak and write only Standard English, in 1987, the State of Hawaii Department of Education acknowledged that Hawaii Creole English constituted a language in its own right with a structure distinct from English. However, this recognition has not resulted in any concrete action in the classroom.

Like most pidgin and creole languages, Hawaii Creole English has not been standardized. This gives rise to the popular belief that the language cannot be written or that it is not a language at all. Linguists have for some time been using a phonemic orthography (i.e. a writing system whose letters closely correspond to the actual speech sounds) developed by Carol Odo (1975). However, it has no wider recognition. Despite the lack of written norms and standardization, there are nevertheless some writers who have attempted to use the creole as a medium for poetry, short stories, and drama by adapting English spelling to represent some of the distinctive characteristics of speech varieties in Hawaii. Each writer has worked out his or her own ad hoc spelling system. This burst of literary creativity can likewise be seen partly as a manifestation of opposition to colonialism and as an affirmation of distinctive local identity in which the use of creole plays a key role.

Although most of the vocabulary of Hawaii Creole English is derived from English, a significant number of words come from Hawaiian. As many as a thousand Hawaiian words may have been in use at one time during the plantation era, of which 250 to 350 were in fairly common use colloquially among the population. Although this number is now fewer, many still persist in local English and many more in the Hawaii Creole English of older speakers. Some words, such as *hemo*, 'take off, remove', and *ho'omalimali,* 'flatter, curry favor', are not as widely known among younger people now, whereas virtually everyone resident locally would know *haole,* 'foreigner'. Because of the general decline in the knowledge of Hawaiian, many people who use a great many of these Hawaiian words every day, either in Hawaii Creole English or in other local forms of English, do not realize that commonly used words such as *akamai,* 'clever, smart', *pio,* 'to turn off, extinguish',

huhu, 'angry, annoyed', and *ono,* 'tasty', are actually Hawaiian, rather than Japanese or even English in origin. Many of these Hawaiian words are being replaced by English ones, although in some cases, the English variants are still different from mainland US English, e.g. *grinds,* 'food' (cf. *kaukau,* 'to eat/food'). The effect of Hawaiian on standard English outside Hawaii is otherwise negligible, consisting mainly of items such as *aloha*, *lei*, *hula*, *ukulele,* and a few others.

Distinctive grammatical features include the lack of the copula *be* (*now you da head man,* 'Now you are the head man'), use of *get* in both possessive and existential constructions (*get one wahine she get one daughter,* 'There is a woman who has a daughter'), use of *stay* for locatives and progressives (*Kimo stay inside da house,* 'Kimo is inside the house', *Kimo stay eating,* 'Kimo is eating'), use of *wen* or *had* as a simple past tense marker (*Charlene wen/had play with da kids,* 'Charlene played with the kids'), use of *pau* (Hawaiian 'done, finished') as a completive marker (*Come back when you pau,* 'Come back when you are finished'), preverbal negation (*Harold no mo rice,* 'Harold doesn't have any more rice'), and use of *for* as a complementizer (*She like know how fo write pidgin,* 'She wants to know how to write pidgin'). The word order is generally subject–verb–object, as in Standard English, apart from topic/comment structures such as *big, da fish,* 'The fish is big', in which the comment appears first. Formerly, there was variation in basic word order among pidgin speakers of different ethnic backgrounds, reflecting the influence of immigrant languages such as Japanese, e.g. *sometime good road get,* 'Sometimes there are good roads'.

There is considerable variation in pronunciation among local residents of different ethnic and social class backgrounds. Generally speaking, however, the phonology of Hawaii Creole English has a smaller inventory of distinctive sounds than many mainland varieties of English. The [r] after a vowel in words such as *shark* is usually absent in the most creole-like speech varieties (i.e. [shak]), and many of the diphthongs (double vowels) found in mainland varieties of English in words such as *boat* [bout] and *play* [pleI] are single vowels in the creole (i.e. [bot], [ple]). Hawaii Creole English often has full vowels where mainland varieties use reduced ones, e.g. creole [tude], 'today', vs. mainland [tədei], and creole [mauntən] 'mountain' vs. mainland [maunten]. English interdental fricatives ('th') in words such as *they* and *think* tend to become stops in the creole (i.e. [de], [tink]). The stops in consonant clusters such as [tr] in words such as *try* are affricated (i.e. [črai]), and initial [s] in clusters such as [str] in words such as *street* sounds more like 'sh' [š]. There are also some stress and intonational differences, such as the use of falling pitch for

yes/no questions, which also do not show the subject/auxiliary inversion typical of standard English, e.g. *you like go Honolulu*? 'Do you want to go to Honolulu?' The falling intonation pattern has been carried over from Hawaiian into creole. Rising pitch together with a final question particle are used as a confirmation check in utterances such as *no mo job fo you, aeh*? 'There isn't a job for you, right?'

References

Bickerton, Derek. 1977. Change and variation in Hawaiian English, Vol. I, Creole Syntax. Honolulu: Social Sciences and Linguistics Institute, University of Hawaii.

Odo, Carol. 1975. Phonological processes in the English dialect of Hawaii, Ph.D. Dissertation. University of Hawaii at Manoa.

Reinecke, John E. 1969. Language and dialect in Hawaii. A sociolinguistic history to 1935. Honolulu: University of Hawaii Press.

Romaine, Suzanne. 1994. Hawaii Creole English as a literary language. Language in Society 23. 527–54.

Romaine, Suzanne. 1999. Changing attitudes towards Hawaii Creole English: Fo' get one good job, you gotta know ho fo' talk like one haole. Creole genesis, attitudes and discourse, ed. by John Rickford and Suzanne Romaine, 287–301. Amsterdam: John Benjamins.

Sato, Charlene J. 1985. Linguistic inequality in Hawaii: the post creole dilemma. Language of inequality, ed. by Nessa Wolfson and Joan Manes. Berlin: Mouton.

Sato, Charlene J. 1991. Language attitudes and sociolinguistic variation in Hawaii. English around the World, ed. by Jenny Cheshire. Cambridge: Cambridge University Press.

Watson-Gegeo, Karen A. 1994. Language and education in Hawaii: sociopolitical and economic implications of Hawaii Creole English. Language and the social construction of identity in creole language situations, ed. by Marcyliena Morgan. Los Angeles: UCLA Center for African American Studies.

SUZANNE ROMAINE

See also **Pidgins and Creoles**

Hebrew: Biblical

Biblical Hebrew (BH) is a Semitic language related to Arabic, Ethiopic (such as Amharic, the national language of Ethiopia), Akkadian (Assyro-Babylonian), etc. It is called 'the language of Canaan' in the Old Testament (Isaiah 19:18). It is also called *yəhūdīt*, 'the language of Judah' or 'Judean', after the dominant tribe in the southern part of the Holy Land (II Kings 18:26, 28). Hebrew is first attested more than 3,000 years ago, and the oldest BH texts date back to the twelfth century BC. In all likelihood, Hebrew became extinct as a spoken tongue in the third century AD; however, in most places it was being supplanted by its kindred language, Aramaic, in the second and first centuries BC. It has been used continuously by Jews around the world as the language of their liturgy and also for literary purposes.

Classification

More specifically, BH belongs to the Canaanite branch of the Northwest Semitic (NWS) group. Many scholars consider Ugaritic to be a member of the Canaanite family as well; however, the present author considers this cuneiform language, discovered in northern Syria in 1929, to be a distinct branch of Central Semitic in its own right (see Kaye 1991). Most authorities would agree that the Canaanite dialects consist of BH,

Phoenician, Ammonite, Edomite, and Moabite. One other language, known as Deir ʿAllā, is controversially considered by some to be Canaanite, but this author does not agree with that classification (see Kaye 1991). The Canaanite dialects are sometimes referred to as languages; however, there existed, in all probability, a high degree of mutual intelligibility among them, along the lines of, say, Egyptian Colloquial, Syrian, Palestinian, and Jordanian Arabic.

The Stages of Biblical Hebrew

Based on linguistic factors that occur in the Hebrew Bible, BH is divided into three different stages: Old BH (1100–1000 BC), Standard BH (1000–550 BC), and Late BH (550–200 BC). The standard language may be exemplified by the Pentateuch. Old BH is best represented by the archaic poetic dialect found in Judges 5 and Exodus 15 (the Song of Moses). The former, known as the Song of Deborah, appears to be Israelian, since it uses a masculine plural in *-īn* for Judahite *-īm*, and the relative particle is *ša-* for *ʾăšer*. Thus, BH definitely had regional dialects: the Judahite dialect (i.e. Jerusalem) and the (northern) Israelian dialect (e.g. Samaria and the Galilee). About four fifths of the Hebrew Old Testament is written in the Judahite dialect. Late BH shows influence from

Aramaic, the *lingua franca* of that era, and is represented by the Books of Esther, Ezra, etc.

Diglossia

BH was a diglossic language. Diglossia is a sociolinguistic situation where a high and a low variety of a language coexist (see Rendsburg 1990, Kaye 1993). Each variety is used under a different set of circumstances, viz. formal or official discourse vs. informal (the closest parallels are the diglossic cases of Arabic and Modern Greek). Examples of colloquial usage have, for a variety of reasons, entered the Biblical text, since the Old Testament should not be regarded as a compilation of speech from Biblical times, but rather as an edited collection of essays and literary versions of direct speech and oral transmissions.

Orthography

BH uses the Phoenician alphabet of 22 letters, including some polyphonous graphemes, and is written from right to left. All traditions of Ancient Hebrew agree that there are 23 consonant phonemes, including two semiconsonants that developed from a system of 29 Proto-Semitic consonants. Originally, only the consonants were represented, and the corpus had achieved its current stable state by the second century AD. As the orthography developed, a system of *matres lectionis* 'mothers of reading', or so-called vowel letters, emerged, viz., <h>, <w>, and <y> were used to indicate their respective phonetically similar corresponding vowels in final position. This system, in existence since the tenth century BC, expanded over time to include medial vowels as well (attested since the ninth century BC), while initial vowels did not need to be included, since no BH word could begin with a vowel (with the exception of /ū/, an allomorph of /wə-/ 'and'). The orthographic representations without *matres lectionis* are older spellings than those with them; <kl> 'each; every' is spelled only once as <kwl> (cf. Arabic <kl>, which is the cognate).

The Hebrew Bible's oral tradition was preserved from the sixth to the tenth century AD by a group of Jewish scholars known as the Masoretes (< Hebrew *māsōrāh* 'tradition'). These scholars established a recognized text of the Hebrew Bible called the Masoretic Text. Three different systems developed: Babylonian, Palestinian, and Tiberian. The authoritative one since about AD 1000 has been the Tiberian, the best known of which is the work of Aaron Ben-Asher (AD 915). A system of diacritics, mostly for vowels, developed, in addition to accent markings, cantillation marks for liturgical recitation, and marginal notes. Its purpose was to insure the accuracy of recording the traditional pronunciation of the Biblical text. Among the most significant of all Masoretic inventions was the dot, called *daghesh*, which was used inside some consonants to indicate a geminated or a stop version of that consonant (as opposed to a fricative, since like Aramaic, BH developed fricative allophones for /bgdkpt/, viz. [βɣðxfθ]). Consonantal gemination was phonemic in BH as it is in Arabic; however, a characteristic of the Masoretic text was that five consonants could not be geminated (probably for phonetic reasons, all being back consonants). They are: /ʾʕhḥr/. Since they do not geminate, vocalic change was commonplace with the preceding vowel via a process of compensatory lengthening.

Basic Characteristics of Biblical Hebrew

BH has two grammatical genders, masculine (*ʾīš* 'man') and feminine (*ʾiššɔ* 'woman'). It also has a dual for nouns that occur in pairs, such as for many body parts (e.g. /ˈēynayim/ 'eyes'). It has a triconsonantal root system like the other Semitic languages (see further below); e.g. the root *qdš* has to do with 'holiness' or 'sanctification'. Furthermore, it is well known for its seven major *binyanim* or verbal forms or stems (Arabic has 10!). Illustrating with the aforementioned root, they are as follows (in the perfect, third-person masculine singular, the usual or unmarked citation form for the BH verb, following Steiner 1992):

I *qal* ('simple' or basic) *qɔdaš* 'be holy'
II *nifʕal* (passive of I) *niqdaš* 'reveal oneself as holy'
III *piʕel* (intensive or causative) *qiddeš* 'sanctify; purify'
IV *puʕal* (passive of III) *quddaš*
V *hifʕīl* (causative) *hiqdīš* 'consecrate; devote'
VI *hufʕal* (passive of V) *huqdaš* 'be consecrated; devoted'
VII *hitpaʕel* (reflexive-reciprocal) *hitqaddeš* 'sanctify or purify oneself'

There are also minor stems, such as another causative *šafʕel*, which derives from a Proto-Semitic and Proto-Afroasiatic *s*-causative. The root *ktb* 'write' may serve as illustrious of nominal and verbal morphology (it occurs 203 times in the Old Testament). Among the verbal forms to be noted are *kɔtab* 'he wrote', *kɔtūb* 'written', *niktab* 'it was written', *məkatbīm* 'write constantly (masc. pl.)', and the nominal patterns *kətɔb* 'register', *kətobet* 'tattooing', *miktab* 'writing', and, possibly, *miktɔm* 'psalm of expiation' via *b* > *m*.

Further, BH has a perfect and an imperfect, which eventually came to denote the past and the present-future (see below).

Phonology

Among the most interesting of the BH phonological developments from Proto-Semitic are the changes of the interdentals to corresponding fricatives, creating mergers; e.g. */ð/ and */z/ > /z/, and */š/ and */θ/ > /š/. Also noteworthy are the following: */ṣ/, */ṣ́/, and */θ̣/ > /ṣ/, */ɣ/ and */ʿ/ > /ʿ/, and */x/ and */ħ/ > /ħ/. The best-known BH phonological fact, in all likelihood, has to do with a dialectological difference reported in the Book of Judges, where the Ephraimites were unable to pronounce *shibboleth* 'flowing stream' or 'ear of corn' correctly as in the Gileadite dialect, saying *sibboleth* instead (Judges 12:6), and this erroneous pronunciation gave away their identities. It should be pointed out that the letters *šīn* and *sīn* are written with the same grapheme and thus were polyphonous, since /ś/ merged with /š/ over time. One also notes several dialectal variants with sibilants, e.g. *śħq* and *ṣħq* 'laugh', and *zʿq* and *ṣʿq* 'shout'.

Emphatics

The exact articulation of emphatic phonemes /ṣ/, /ṭ/, and /q/ are not really known. The evidence from the cognate languages is mixed. If one compares with Arabic, the corresponding sounds are pharyngealized and velarized, except for the voiceless uvular plosive. On the other hand, if one adverts to Ethiopic and South Arabian, the sounds are glottalized (ejective). Most Semitists believe that the evidence favors the glottalization of the Proto-Semitic emphatics.

Morphology

We have already mentioned the triconsonantal root (see above). However, it is important to keep in mind that early Semitic also had many biconsonantal roots. Unquestionably, many original biconsonantal roots became triconsonantal via *Systemzwang*. Basically, the idea is that a root determinative was added to make an original biconsonantal triconsonantal. The details of this process are not fully understood, and this area is rife with controversy.

Perhaps the most controversial area in BH morphology has to do with the nature of the verbal system (see above). Many conclude that the imperfect originally expressed actions and was used to denote the durative aspect. The perfect originally expressed the stative. The preterite nature of the imperfect is preserved in the BH 'consecutive' or 'conversive' tenses. Many of the 5,280 occurrences of the root *ʾmr* 'say' are with the *wāw*-consecutive: *wayyōmɛ/ar* 'and he said'. As is the case in BH, the *wāw*-consecutive form also occurs in Ugaritic prose, not poetry.

Syntax

The adjective follows the noun and agrees with it in number, gender, and definiteness. Thus, *ʾiššɔ kūšīt* 'an Ethiopian woman' and *hɔʾiššɔ hakkūšīt* is 'the Ethiopian woman'.

The basic word order is VSO, but for emphasis other word orders are possible. There is an accusative particle *ʾet* that marks definite direct objects. Thus, Genesis 1:1 reads: *bərēšīt bɔrɔ ʾĕ/ōhīm ʾet haššɔmayim wəʾet hɔʾɔrɛṣ,* lit., 'in the beginning-created-God-Acc.-the heavens-and-ACC.-the earth'.

Loanwords

BH has borrowed vocabulary from a variety of languages, including Egyptian, Sumerian, Akkadian, Aramaic, Greek, and Persian. Two well-known examples include *barzel* 'iron' < Hittite *barzillu*, and *parʿō* 'Pharaoh' < Ancient Egyptian *pr-ʿȝ* 'the great house'. Examples of BH words that have come down into English include many proper names (Saul, David, etc.) and terms relating to religion, such as *amen* and *sabbath*, not to mention the ubiquitous *schwa*, the mid-central vowel [ə] known to all linguists.

References

Blau, Joshua. 1976. A grammar of Biblical Hebrew. Wiesbaden: Otto Harrassowitz Verlag.

Gesenius, Wilhelm, and E. Kautsch. 1910. Gesenius' Hebrew grammar, 2nd English edition. Oxford: Clarendon Press.

Kaye, Alan S. 1991. Does Ugaritic go with Arabic in Semitic genealogical sub-classification? Folia Orientalia. Vol. 28, pp. 115–28.

Kaye, Alan S. 1993. Review of Rendsburg 1990. AJS Review.

Kaye, Alan S. (ed.) 1997. Phonologies of Asia and Africa. Winona Lake, Indiana: Eisenbrauns.

Khan, Geoffrey. 1997. Tiberian Hebrew phonology. In Kaye.

Koehler, L., W. Baumgartner, and J. J. Stamm. 1967–1990. Hebräisches und Aramäisches Lexicon zum Alten Testament, 3rd edition. Leiden: E. J. Brill.

Kutscher, E. Y. 1982. A history of the Hebrew language, ed. by Raphael Kutscher. Jerusalem: Magnes Press and Leiden: E. J. Brill.

Malone, Jopseph L. 1991. Tiberian Hebrew phonology. Winona Lake, Indiana: Eisenbrauns.

Rendsburg, Gary A. 1990. Diglossia in ancient Hebrew. New Haven, CT: American Oriental Series No. 72.

Rendsburg, Gary A. 1997. Ancient Hebrew phonology. In Kaye.

Sáenz-Badillos, Angel. 1993. A history of the Hebrew language. Cambridge: Cambridge University Press.

Steiner, Richard C. 1992. Ancient Hebrew. In International encyclopedia of linguistics, ed. by William Bright.

Ullendorff, Edward. 1977. Is Biblical Hebrew a language? Wiesbaden: Otto Harrassowitz Verlag.

ALAN S. KAYE

See also **Hebrew: Modern ; Semitic Languages**

Hebrew: Modern

Modern Hebrew (MH) is the first language for the native Jewish population in the State of Israel and the second language for the new immigrants and the native Arab population. In its current form, MH is dated back only as far as the late nineteenth century; prior to that period, Hebrew was used for liturgical and ritual purposes for about 1,700 years. The genetic affiliation of MH to the Semitic family, although not uncontroversial (Wexler 1990), is strongly manifested in its morphology and lexical stock, which were drawn from Biblical Hebrew, the language of the Old Testament. The phonology of the language, on the other hand, as well as word order, do not reflect the characteristics of Biblical Hebrew.

Verb Morphology

MH morphology is characterized by the Semitic type nonconcatenative structure, especially in the verbal system. A verb must belong to one of the five morphological classes called *binyanim* 'constructions' (singular *binyan*). Each *binyan* has a specific prosodic structure (in terms of syllables) and vocalic pattern, and some also have a prefix; for some *binyanim,* these properties must be specified for each tense. The verbs below are given in their third-person singular form; stress is final unless otherwise specified (by an acute sign). (see Table 1)

While the verbs in table 1 differ in their *binyan* (i.e. prosodic structure, vocalic pattern, and prefixes, if any), they all share a discontinuous string of consonants, traditionally called the 'consonantal root'. Verbs sharing a consonantal root often share a semantic property, and it is thus traditionally claimed (for all Semitic languages displaying this type of morphology) that the consonantal root carries the core meaning of the verb. Notice that not every root has a verb in all the *binyanim,* as can be seen from the different B2 verb in the table in table 1, which replaces the

nonexistent verb **nigdal*. It appears that there is no other language family that is claimed to have roots consisting of consonants only. For this and other reasons, the validity of the notion of the root has been challenged (Bat-El 2002). The *binyan* may contribute a syntactic property to the verb, but only in paradigmatic relations (Berman 1978). For example, B2 is the passive counterpart of B1 (*gamar* 'finish B1'—*nigmar* 'be finished B2'), B4 is the transitive counterpart of B1 (*lamad* 'learn B1'—*limed* 'teach B4'), B3 is the causative counterpart of B1 (*xatam* 'sign B1'—*hextim* 'have someone sign B3'), and B5 is the reciprocal, reflexive, or passive counterpart of B4 or B1 (*niʃek* 'kissed B4'—*hitnaʃek* 'kissed with someone B5', *raxac* 'wash B1'—*hitraxec* 'wash oneself B5', *kibel* 'receive B4'—*hitkabel* 'be received B5'). There are two passive forms, often considered to be *binyanim* as well, derived from B3 and B4 by substituting the vocalic pattern with {u, a} (B3 *higdil* 'enlarge'—*hugdal* 'be enlarged', B4 *gidel* 'raise'—*gudal* 'be raised'). These forms, although not commonly used, have no exceptions and are always contingent, structurally and semantically, upon their active counterpart.

Verbs are also accompanied by affixes indicating tense, person, number, and gender. (see Table 2)

The values of the person–number–gender properties of the verb are drawn from subject noun or pronoun with which the verb must agree (*ha-yalda gadl-a* 'the girl grew', *ha-yeladim gadl-u* 'the children grew', *ʔanaxu gadal-nu* 'we grew'). The gender properties of the noun are natural in case of animate nouns (*sus* 'horse ms.' - *susa* 'horse fm.'), but serve in a purely grammatical role in inanimate nouns (*séfer* 'book ms.', *délet* 'door fm.'). Subject–verb agreement allows phrases without a subject, but only when the subject is not the third person (*karánu séfer* '(we) read a book'). There are also phrases without a verb, but only in the present tense (*ha-séfer al ha-madaf* 'the book (is) on the shelf', *ha-yalda xaxama* 'the girl (is) smart').

TABLE 1

Binyan/Tense	Past	Participle	Future	
B1 paʔal	gadal	gadel	yigdal	'grow'
B2 nifʔal	(nigmar)	(nigmar)	(yigamer)	('be finished')
B3 hifʔil	higdil	magdil	yagdil	'enlarge'
B4 piʔel	gidel	megadel	yegadel	'raise'
B5 hitpʔael	hitgadel	mitgadel	yitgadel	'agrandize'

TABLE 2

Gender/Person	Past		Future		Participle		
	Singular	Plural	Singular	Plural		Singular	Plural
M/F - 1	gidál-ti	gidál-nu	ʔa-gadel	ne-gadel			
M - 2	gidál-ta	gidál-tem	te-gadel	te-gadl-u	M	me-gadel	me-gadl-im
M - 3	gidel	gidl-u	ye-gadel	ye-gadl-u			
F - 2	gidál-t	gidál-tem/n	te-gadl-i	te-gadl-u			
F - 3	gidl-a	gidl-u	te-gadel	ye-gadl-u	F	me-gadél-et	me-gadl-ot

Noun Morphology

Nouns, like verbs, exhibit prosodic and vocalic restrictions called *mishkal* 'weight' (plural *mishkalim*). Related nouns (and adjectives) differ in their prosodic structure and vocalic pattern, as well as affixation (*gadal* 'to grow'—*gódel* 'size', *mi-gdal* 'tower', *gadol* 'big'; *carax* 'to consume'—*córex* 'need', *ti-cróx-et* 'consumption', *carx-an* 'consumer'). Not only are there far more *mishkalim* than *binyanim*, a noun, unlike a verb, does not have to fit into a *mishkal*. Thus, while loan-verbs (or verbs derived from loan nouns) must belong to a *binyan* (*tilfen* 'to phone', *kided* 'to codify'), nouns can maintain the structure of the source word, with a few phonological modifications (*diéta* 'diet', *télefon* 'phone'). The *mishkal* system is not rigid because there is nothing in the morphology of nouns that is contingent upon the *mishkal*. In verbs, in order to be able to derive one tense from the other, it is crucial to know the *binyan*, which is identified on the basis of the prosodic structure and vocalic pattern (and prefixes, if any). Such information is not relevant for the morphology of nouns, which involves mainly suffixation: plural suffixes *-im* for masculine (*xatul-im* 'cats ms.') and *-ot* for feminine (*xatul-ot* 'cats fm.'), various feminine suffixes (*xatul-a* 'cat fm.', *rakdan-it* 'dancer fm.', *sofér-et* 'writer fm.'), a diminutive suffix (*sus-on* 'little horse'), an agent suffix (*taklit* 'record'—*taklit-an* 'DJ'), various nominal suffixes (*manhig* 'leader'—*manhig-ut* 'leadership', *mila* 'word'—*mil-on* 'dictionary', *magav* 'wiper'—*magév-et* 'towel'), and an adjectival suffix (*tarbut* 'culture'—*tarbut-i* 'cultural'). The suffixes *-a*, *-it*, *-et*, and *-ut* assign a feminine gender and all the others assign masculine gender to the base. The plural suffixes do not assign gender but rather agree with the gender of the base (with several exceptions, such as *ʃoʃana* 'rose fm', which takes the masculine plural suffix *-im*, *ʃoʃan-im*). These suffixes are not sensitive to the structure of the noun to which they are attached and one can even find borrowed nouns with a native suffix (*geográf-i* 'geographic', *dialóg-im* 'dialogs', *simétriy-ut* 'symmetry'). The participle form of the verb also functions as a noun, and its morphology is that of the noun, i.e. it has only number and gender specification (*ha-ʃomer ʃomer al ha-báyit* 'the guard guards (ms.sg.) the house', *ha-ʃomer-im ʃomer-im al ha-báyit* 'the guards guard (ms.pl.) the house').

Beyond the Word

A noun can also be formed from two nouns (and an adjective from an adjective plus a noun), where the first element determines the general meaning, thus considered the head, while the second adds specific information. For example, *beit+séfer* 'school (house+book)' is a type of a house, *naʔarat+rexov* 'street girl (girl+street)' is a type of a girl, and *ʔiver+cvaim* 'color blind (blind+colors)' refers to a type of blindness. There are two types of noun+noun structure, one called a compound and the other a construct state, where only the latter is accessible to syntactic modification (Borer 1989). In the construct state, the non-head (second) element can be modified (*gan+ʃoʃanim ʔadumot* 'a garden of red roses (garden+roses red)'), or conjoined with another noun (*gan+ʃoʃanim ve-cipornim* 'garden of roses and carnations (garden+roses and carnations)'). However, in a compound such as *gan+yeladim* 'kindergarten (garden-children)' the non-head *yeladim* is not accessible to the syntax. Despite the difference in accessibility to syntax, both the compound and the construct state display the same morphology and phonology. In both, the plural suffix, which has the special form *-ei* for masculine, is attached to the first (head) element (*gan-ei +yeladim*, *gan-ei+ʃoʃanim*), the feminine suffix *-a* takes the form *-at* (*naʔara* 'girl' - *naʔarat+rexov* 'street girl'), and, in both, the second element bears the main stress. Occasionally, speakers may disagree as to whether the noun+noun form is accessible to the syntax, i.e. whether it is a compound or a construct state.

Word order in MH is usually subject–verb–object (SVO), but permutation of this order is relatively free. Thus, the phrase *dáni raʔa ʃoʃana ʔaduma* 'Dani saw a red rose' can appear in some contexts as *ʃoʃana ʔaduma dáni raʔa* or *ʃoʃana ʔaduma raʔa dáni*. However, adjectives always follow the noun they modify (*ʃoʃana ʔaduma* 'red rose (rose red)'), and numerals, with the exception of the numeral 'one', always precede it (*ʔéser ʃoʃanim* 'ten roses', *ʃoʃana ʔaxat* 'one rose (rose one)'). Definite nouns are preceded by the definite article *ha-* 'the', which also appears in the

modifying adjective (**ha**-ʃoʃana **ha**-ʔaduma); there is no indefinite article. All other function words also appear as proclitics: prepositions appear before the phrase (**ʔal**-ha-séfer 'on the book', **me**-ha-báyit 'from the house'), the conjunctive marker ve- 'and' appears before the last element in the list (séfer, niyar **ve**-ʔet 'a book, a paper, and a pen'), and the subordination marker ʃe- 'that' appears before the subordinate clause (ha-baxur **ʃe**-xika la 'the guy that waited for her'). Question words (mi 'who', ma 'what', matay 'when', ʔeyfo 'where') appear at the beginning of the phrase (**mi** zarak kadur 'who threw a ball?') and can be preceded by a preposition (**ʔal-mi** zarákta kadur 'at whom did you throw a ball?').

Phonology

The sound system of MH consists of 20 consonants— *p, b, m, f, v, t, d, s, z, c, n, l, ʃ, y, k, g, x, r, h, and ʔ* (and in some dialects also ʕ and ħ), and five vowels; *i, u, e, o, a*. The sound *h*, is often substituted by *ʔ*, and both are deleted in casual speech.

Stress in nouns can be ultimate or penultimate (and rarely antepenultimate). When the plural suffix is added, stress shifts to the end in most nouns, whether the singular form has ultimate stress (gamád - gamadím 'dwarf sg.-pl.'), or penultimate (yéled - yeladím 'boy sg.-pl.'); in some nouns, the stress stays in the same position on the stem (salát - salátim 'salad', tíras - tírasim 'corn sg.-pl.'). In verbs, stress is always final on the stem (gadál 'he grew', gadál-nu 'we grew'), but shifts to the end when a vowel initial suffix is added (gadl-á 'she grew'). When stress shifts, the vowel in the stem final syllable is deleted when the preceding syllable is open (gadal-a —-> gadlá 'she grew'), and changes to *e* when the preceding syllable is closed (yigdal-u —-> yigdelú 'they will grow') or when flanked by identical consonants (garar-a —-> garerá 'she dragged'). Stress shift and vowel deletion/change do not affect B3 verbs (higdíl-a 'she enlarged') and

monosyllabic verbs (rác-u 'they ran'). In nouns, it is the vowel in the penultimate stem syllable that undergoes deletion, but only in a lexically marked group of nouns and adjectives (gamal - gmalim 'camel sg.-pl.' but gamad - gamadim 'dwarf sg.-pl.'). When the first consonant in the penultimate syllable is a sonorant (*m, n, l, r, y*), the vowel subject to deletion is replaced by *e* (lavan - levanim 'white sg.-pl.').

There is a stop-fricative alternation in MH (*p - f, b - v, k -x*), reminiscent of the postvocalic spirantization of Tiberain Hebrew (**k**atav 'he wrote' - yi**x**tov 'he will write'). However, this alternation is not entirely regular as there are postvocalic stops (bike ʃ 'he requested') as well as fricatives in non-post-vocalic positions (**x**ipes 'he searched'). At the current stage of the language, there is a great degree of free variation in the verb system (yixtov ~ yiktov), which may suggest a change toward a uniform verbal paradigm with respect to spirantization.

References

Bat-El, Outi. 2002. Semitic verb structure within a universal perspective. Language processing and acquisition in languages of Semitic, root-based, morphology, ed. by Joseph Shimron. Amsterdam: John Benjamins.

Berman, Ruth A. 1978. Modern Hebrew structure. Tel-Aviv: University Publishing Projects, Ltd.

Borer, Hagit. 1989. On the morphological parallelism between compounds and constructs. Morphology yearbook 1, ed. by Geert Booij and Jaap van Marle. Dordrecht: Foris.

Glinert, Lewis. 1989. The grammar of modern Hebrew. Cambridge: Cambridge University Press.

Ravid, Dorit D. 1995. Language change in child and adult Hebrew: a psycholinguistic perspective. New York: Oxford University Press.

Rosen, Haiim B. 1977. Contemporary Hebrew. The Hague: Mouton.

Wexler, Paul. 1990. The schizoid nature of modern Hebrew: a slavic language in search of a semitic past. Wiesbaden: Otto Harrassowitz.

OUTI BAT-EL

See also **Hebrew: Biblical ; Semitic Languages**

Hindi-Urdu

Hindi is a modern Indo-Aryan language spoken in South Asian countries (India, Pakistan, and Nepal) and also in countries outside Asia (Mauritius, Trinidad, Fiji, Suriname, Guyana, South Africa, and other countries). Approximately six to seven hundred million

people speak Hindi either as a first or second language. It is ranked among the two most widely spoken languages of the world. Along with English, it is the official language of India. In addition, it is the state language of Bihar, Haryana, Chattisgarh, Jharkhand,

Himachal Pradesh, Arunachal Pradesh, Madhya Pradesh, Uttar Pradesh, and Rajasthan.

Hindi, which is a descendant of the Sanskrit language, is not strictly the name of any chief dialect of the area but is an adjective, Persian in origin, meaning Indian. Historically, it was synonymous with Hindui, Hindawi, Rexta, and Rexti. The term Urdu is also used to refer to this language. All these labels denote a mixed speech spoken around areas of Delhi, North India, which gained currency during the twelfth and thirteen centuries as a contact language between Arabs, Afghans, Persian, Turks, and native residents. Hindi and Urdu have a common form known as Hindustani. This was the variety that was adopted by Mahatma Gandhi and the Indian National Congress as a symbol of national identity during the struggle for freedom from Britain.

Urdu is spoken by approximately 50 million speakers in the subcontinent of India and Pakistan. It is the national language of Pakistan and one of the 19 official languages of India. The term 'Urdu' is Turkish in origin (i.e. ordu 'camp'). As mentioned above, both Hindi and Urdu have their origins in the mixed speech spoken around Delhi. In time, this mixed speech called khari boli developed its own variety called 'Hindi' or 'Urdu'.

The first question that is immediately raised about Urdu is: How does it differ from Hindi? This is not an easy question to answer. In many respects, the two speech varieties are mutually intelligible and at another level they are quite unintelligible. Moslems often report their speech variety as Urdu, and Hindus and other non-Moslem speakers tend to report their speech variety as Hindi. Yet, one notable difference in Hindi and Urdu is script.

Hindi is written in the Devanagari script, which is ranked as the most scientific writing system among the existing writing systems of the world. The Devanagari script is written from left to right and is a descendant of the Brahmi script, which was well established in India before 500 BC. The script is phonetic in nature and there is a fairly regular correspondence between the letters and their pronunciation.

Urdu is written in the Persio-Arabic script. Like Arabic, it is written from right to left and short vowels are usually not written. It is due to Urdu that the Arabic script underwent great modification and some new symbols were added to it to represent sounds, such as retroflex consonants (produced with the tip of the tongue curled upwards), which are specific to the Indian subcontinent.

In addition to Hindi and Urdu being written in two different scripts, the two languages also differ from each other in minor ways in their grammatical systems and vocabulary. Urdu tends to borrow its vocabulary from Persio-Arabic sources, whereas Hindi borrows from indigenous Sanskrit sources. This tendency has resulted in the admission of new sounds (e.g. z, x, G) into Urdu and in the development of its distinct literary style.

The literary history of Hindi goes back to the twelfth century. Some notable literary figures of Hindi are Surdas, Tulsidas, and Mira Bai. Urdu literature flourished both in India and Pakistan. The literary history of Urdu can be traced back to approximately the thirteenth century. Its first poet was Amir Khusro (1253–1325), who termed the language *zabaan-e-Dehlvi* or Hindi. Amir Khusro included Hindi verses among his Persian poems and romances. But the greatest poet of Hindi-Urdu was Kabir, who was famous for his devotional poetry.

Urdu is particularly well known for its romantic literature. The two most famous genres of Urdu are Masnawii and Gazal, which are the gift of the mixing of the Hindu and Persian-Muslim cultures. Some notable literary figures of Hindi-Urdu literature are Inshah Alla Khan, Malik Muhammad Jayasi, Mir, Daya Shankar Nasim, Bahadur Shah Zafar, Iqbal, and Mirza Ghalib.

Hindi-Urdu has an approximately three-century-old, well-attested, and rich grammatical tradition of its own. It is a by-product of the colonial era and was born shortly after the arrival of Europeans in India.

Phonology

The two notable phonological features of the language are as follows: (1) Hindi still retains the original Indo-European (1500 BC) distinction between aspirated and unaspirated consonants, which results in a four-way contrast as shown in the following examples: *kaal* 'time', *khaal* 'skin', *gaal* 'cheek', and *ghaal* 'to put into'. (2) It has the feature of retroflexion in its consonant inventory, cf. *Taal* 'to put off' and *taal* 'pond'. The retroflex consonant is transcribed as the capital T.

The inventory of distinctive segments of standard Hindi-Urdu is shown in Table 1. Sounds that occur in Persio-Arabic words are provided in parentheses ( ).

Stress

Although stress (meaning loudness) is not a prominent feature of Hindi-Urdu, nevertheless, it seems that its existence cannot be denied. Stress can distinguish between grammatical categories such as nouns and verbs, as in

Nouns	Verbs
galaa neck	ga**laa** cause to melt
talaa sole	ta**laa** cause to fry

TABLE 1 Consonants

		Labial	Dental	Retroflex	Palatal	Velar	BackVelar
Stop	Unvoiced unaspirate	p	t	T	c	k	(q)
	Unvoiced aspirate	ph	th	Th	ch	kh	
	Voiced unaspirate	b	d	D	j	g	(G)
	Voiced aspirate	bh	dh	Dh	jh	gh	
	Nasal	m	n	N	ñ	ŋ	
Fricative	Unvoiced	(f)	s	sh		(x)	
	Voiced		(z)				
Flap	Voiced unaspirate		r	R			
	Voiced aspirate			Rh			
Lateral			l				
Semivowels			w (v)		y		

Vowels

	Front	Central	Back
High	ii		uu
	i		u
Mid	e		o
	ai	a [schwa]	au
Low		aa	

The stressed syllable is shown in bold. However, stress is not usually distinctive in Hindi-Urdu. Therefore, in general, whether one places stress on the first syllable or on the second, the meaning will not be affected. For example, the meaning of word, *sunaa* heard will remain unchanged whether one places stress on the first syllable or the second. Therefore, Hindi-Urdu is often characterized as a 'syllable-timed' language like French, where the syllables are pronounced in a steady flow, resulting in a 'machine-gun' effect.

Morphology

Word formation in Hindi-Urdu primarily uses prefixes and suffixes to define inflectional and derivational word classes. Nouns are generally inflected for number, gender, and case. There are two numbers, singular and plural; two genders, masculine and feminine; and three cases, simple, oblique, and vocative. The oblique forms occur when a noun or noun phrase is followed by a postposition, vocative case marks nouns referring to the addressee of the utterance. Nouns are inflected according to their gender and word-final sound, as exemplified by the four paradigms given in Table 2.

Adjectives are primarily of three types: (1) Simple adjective, such as *acchaa* 'good'; (2) Derived adjectives using various parts of speech such as nouns: *mardaanaa* 'masculine', adverbs: *mandaa* 'slow', and from agentive/adjectival particle *vaalaa*, e.g. *dillii vaalaa* 'from Delhi'; and (3) Participial adjectives: *caltii* 'moving', *bhaagtaa* 'running'. Adjectives can be used both attributively (immediately placed before nouns) and predicatively (immediately placed before verbs). Simple, participial, and *vaalaa* adjectives are of two types: inflected and uninflected. Inflected adjectives agree with their following noun in number, gender, and case; they end in morpheme -aa (e.g. *acchaa* 'good'), which changes to e for masculine plural and masculine oblique (*acche*), ii for feminine nouns (*acchii*). Uniflected adjectives remain unchanged.

Although the case system of pronouns is essentially the same as that of nouns, pronouns have more case forms than nouns. Case relations are essentially carried out by means of postpositions. Personal pronouns are similar to their English equivalents except that there are no gender distinctions (like *he* and *she* in English) (Table 3).

Adverbs and postpositions are invariant, except for the genitive postposition, which behaves like an inflected adjective. The postpositions mark case relations and adverbial functions.

Hindi has no articles.

Verbs

There are three tenses in Hindi: present, past, and future. The tenses are formed by the suffixation process. Verbs are inflected for number, gender, and person.

(1) vo aa-t-aa hai.
    he come-present-mas.Sg is (3sg)
    'He comes'.

(2) vo aa-yaa.
    he come-past-mas, Sg.
    'He came'.

(3) vo aa-ye-g-aa
    he come-3sg-future-mas. Sg
    'He will come'.

TABLE 2   Hindi Noun Paradigms

Case	Paradigm-I Masculine Nouns Ending in *-aa*		Paradigm-II Masculine Nouns Not Ending in *-aa*	
	Singular	Plural	Singular	Plural
Direct	*laRkaa 'boy'*	*laRke*	*aadmii 'man'*	*aadmii*
Oblique	*laRke*	*laRkõ*	*aadmii*	*aadmiõ*
Vocative	*laRke*	*laRko*	*aadmii*	*aadmio*
Case	Paradigm-III Feminine Nouns Ending in *-ii*		Paradigm-IV Feminine Nouns Not Ending in *-ii*	
	Singular	Plural	Singular	Plural
Direct	*laRkii 'girl'*	*laRkiyãã*	*maataa 'mother'*	*maataaẽ*
Oblique	*laRkii*	*laRkiyõ*	*maataa*	*maataaõ*
Vocative	*laRkii*	*laRkio*	*maataa*	*maataao*

TABLE 3   Hindi Pronoun Paradigms

	1st Person		2nd Person			3rd Person	
	Singular	Plural	Singular	Plural	Honorific	Singular	Plural
Direct	mãĩ	ham	tuu	tum	aap	vo	ve
Oblique	mujh	ham	tujh	tum	aap	us	un
Genitive	meraa	hamaaraa	teraa	tumhaaraa	aapkaa	uskaa	unkaa

In addition to simple verbs, Hindi has two categories termed 'conjunct' and 'complex' verbs. The class of conjunct verbs is usually derived adding *karnaa* 'to do' or *honaa* 'to be' to noun, adjective, pronoun, or adverb; e.g.

(4) *kaam* 'work', *kaam karnaa* 'to work',
(5) *acchaa* 'good', *acchaa honaa* 'to recover',
(6) *dhiire* 'slow', *dhiire karnaa* 'to slow down',
(7) complex verb: *likh* 'write', *lenaa* 'to take' -→
     likh lenaa 'to write' (for one's own benefit)'.

Hindi is also sensitive to stative/active and volitional/nonvolitional distinction; these four types of distinction are denoted by morphologically related verbs:

(8) *khulnaa* 'to be opened', *kholnaa* 'to open',
(9) *TuuTnaa* 'to be broken', *toRnaa* 'to break'.

Compounding is an integral and very productive process of word formation in Hindi. The noun–noun compounding involves 12 types of compounding. For example, *khaanaa* 'eating' and *piinaa* 'drinking' can be compounded into *khaan–paan*, which means 'life style'.

From the viewpoint of morphological complexity, Hindi can be classified as an agglutinating language. This means that derivation of words takes place by the addition of suffixes to simple or derived stems of major word classes. The process of prefixation is almost exclusively used with nouns and verbs; other word classes rarely participate in this process. The process of suffixation is productive equally with both nouns and verbs.

Syntax

Hindi is a Subject–Object–Verb (SOV) language with a relatively fixed word order. Interrogative or other sentence types do not introduce any changes in word order. In topicalization and focus structure, however, phrases may occur in a marked (i.e. exceptional) position, usually initial. The verb generally agrees with the subject. In transitive perfective sentences, the subject is marked with the *ne* postposition, and the verb agrees with the direct object. A rule of thumb is that the verb never agrees with any constituent that is marked with a postposition.

Any sentence can be negativized by placing the negative particle *nahii* 'not' in the preverbal position.

(10) vo   nahĩĩ   aa-ye-g-aa
     he   not   come-3sg-future-mas. Sg
     'He will come.'

References

Beg, M.K.A. 1988. Urdu grammar: history and structure. New Delhi: Bahari.

Bhatia, Tej K. 1987. A history of the Hindi grammatical tradition. Leiden: E. J. Brill.

Bhatia, Tej K. 1996. Colloquial Hindi. London: Routledge.

Bhatia, Tej K., and Ashok Koul. 2000. Colloquial Urdu. London: Routledge.

Dimock, Edward C., Edwin Gerow, C.M. Naim, A.K. Ramanujan, Gordon Roadarmel, and J.A.B. van Buitenen. 1978. The literatures of India: an introduction. Chicago: University of Chicago Press.

Kachru, Yamuna. 1980. Aspects of Hind grammar. New Delhi: Manohar Publications.

Ohala, Manjari. 1983. Aspects of Hindi phonology. Delhi: Motilal Banarasidass.

Rai, Amrit. 1984. A house divided. Delhi: Oxford University Press.

Shackle, Christopher, and Rupert Snell. 1990. Hindi and Urdu since 1800: a common reader. New Delhi: Heritage.

Tej K. Bhatia

See also **India**

Hiri Motu

Hiri Motu is a pidgin language that is of particular interest because it is one of the few well-described pidgins with a vocabulary derived from a non-Indo-European language. Its lexical source is the Motu language spoken in a dozen villages along the southern coast of the 'tail' of Papua around Port Moresby, the capital of Papua New Guinea.

Motu is a member of the huge Austronesian language family, which extends throughout Polynesia, Micronesia, much of Melanesia, most of Indonesia and Malaysia, all of the Philippines, Madagascar, and the interior of Taiwan. The Oceanic subgroup, to which Motu belongs, comprises approximately half the total number of Austronesian languages, all of which are located east of a north–south line through the Indonesian province of Papua.

Motu exhibits many of the structural features that are widely distributed in Oceanic languages. There is a set of inflectional prefixes on the verb that provide grammatical information about the subject, with the result that the subject can go unmentioned. These prefixes also mark tense and negation. Verbs that take objects are also marked by suffixes providing grammatical information about the object. Thus, *e-ita-gu* '(s)he saw me' means literally '(s)he:NONFUTURE-see-me' or *baine-ita-gu* '(s)he will see me' (literally '(s)he:FUTURE-see-me').

Although subject and object pronouns are not required in Motu, they can be present. When both subject and object are present, the subject must be followed by *ese*, as in *ia ese lau e-ita-gu* '(s)he saw me' (literally '(s)he SUBJECT me (s)he:NONFUTURE-see-me').

Motu nouns also reflect the widespread Oceanic pattern of distinguishing between inalienable and alienable possession. 'Inalienable possession' refers to an inherent relationship between the possessor and the possessed (e.g. persons and their body parts). In Motu, it is expressed with pronominal suffixes attached directly to the noun, e.g. *mata-gu* 'my eye' (literally 'eye-me'), *mata-mu* 'your eye' (literally 'eye-you'). Alienable possession, i.e. a possessive relationship that is not inherent in the nature of the items concerned, is expressed by means of pronominal suffixes attached to an independent classifying element. The alienable possession of something to be eaten involves the possessive classifier *a-*, while the alienable possession of anything else uses the possessive form *e-*, e.g. *a-gu aniani* 'my food' (literally 'foodstuff-me food'), *e-gu ruma* 'my house' (literally 'thing-me house').

Many Oceanic languages exhibit a word order of S(ubject)-O(bject)-V(erb), while others have SOV. A few languages also have VSO and VOS. Motu falls into the SOV class. SVO languages typically use prepositions to mark certain functions of a noun in the sentence, e.g. the English preposition *at* marks a location. Motu uses postpositions instead (i.e. particles that are equivalent to prepositions but occur after the noun they modify). Thus, *Guhi amo e-moru* '(S)he fell **from** the roof' (literally 'roof from (s)he:NONFUTURE-fall').

With this brief background on Motu, we can turn to Hiri Motu. Because it is a pidgin language, it would be expected to be simplified by comparison with its major source language. Pidgin languages typically eliminate (or sharply reduce) grammatical inflections. This tendency is clearly evident in Hiri Motu in that none of the inflectional prefixes, e.g. Motu's *e-* '(s)he:NONFUTURE', have been retained. Not surprisingly, the independent subject pronouns are obligatory in Hiri Motu rather than optional as in Motu. Tense and negation are marked not by prefixes but by free forms. For example, *do(hore)*, occurring before the verb, marks future tense, while *lasi* 'no' occurs after the verb to mark negation.

The object suffixes have also been lost. Pronominal objects are expressed with the same free pronouns that are found in subject position. Subject and object pronouns are thus distinguished only by word order. With pronominal subjects and objects, the word order is OSV. Thus, contrast the examples above from Motu with the following Hiri Motu equivalents: *lau ia itaia* '(s)he sees/saw me' (literally 'me (s)he see') and *lau dohore ia itaia* '(s)he will see me' (literally 'me FUTURE (s)he see').

459

With noun subjects and objects, the word order in Hiri Motu is SOV, as in Motu. For a pidgin language, this is relatively rare, and even in Pidgin Fijian, for which the lexical source is predominantly OSV, the word order is SVO.

Other features have also been simplified in Hiri Motu. The distinction between inalienable and alienable possession is lost on the grammatical level. However, the distinction is still there on the lexical level: inalienably possessed nouns such as *mata-* 'eye' include the original third-person singular suffix *-na* as part of the noun root. In the grammar, simple possession is marked by a possessive pronoun that is a combination of the independent pronouns and the original nonedible alienable possessive markers. Thus, compare the variety of Motu possessive constructions presented earlier with the following from Hiri Motu: *lauegu matana* 'my eye', *lauegu aniani* 'my food', *lauegu ruma* 'my house'.

Hiri Motu has been discussed so far as a monolithic entity, although there is in fact considerable structural variation. In particular, speakers of Austronesian languages closely related to Motu—predominantly from the coastal area around Port Moresby—incorporate more of the patterns from Motu than do Hiri Motu speakers from farther afield. These people, for example, typically maintain the distinction between inalienable and alienable possession, as well as maintaining object suffixes on transitive verbs.

Papua has long been an area of language contact, and Tom Dutton (1985) describes in detail the development of a number of contact languages from languages of this area in association with the *'hiri'* trading expeditions conducted along the Papuan coast before colonial times. As the *hiri* expeditions were abandoned, so too were the languages that arose in this context, although Dutton was able to record some information from one-time participants.

Pidgin languages based on European languages have typically evolved in language contact situations in plantation economies (whether slave-based as in the Atlantic or contract labor-based as in the Pacific). The development of Hiri Motu is significantly different, in that plantation labor played little role in its initial development. In the Port Moresby area in the 1870s and 1880s, a number of overseas visitors arrived, although, of these, only a relatively small number of Christian missionaries stayed for long periods. The missionaries set out to learn Motu, but many other people did not share the same goal, and a somewhat simplified form of Motu came to be widely used as a lingua franca (and even some of the missionaries ended up learning this simplified variety rather than the variety normally used by Motuan villagers).

In 1888, the British colony of Papua was formally proclaimed. Much of this colony was completely unknown to outsiders. The first colonial administrator immediately set out to establish a police force that would serve to spread and strengthen colonial power. Police constables from a variety of locations were appointed and trained and then posted to different parts of the colony. Some of these constables came from Fiji and the Solomon Islands, where some form of pidgin with an English-based vocabulary was already in use. These initial police trainees were then joined by Papuans from a variety of locations. While it is not known precisely how these constables communicated, it seems possible that they made use of an English-derived pidgin, along with the kind of simplified Motu that they would have acquired as a result of contact with the local Motu-speaking population (with these two varieties possibly influencing each other).

As these constables were posted to outlying areas in the late 1800s and into the 1900s, they took with them a knowledge of whatever variety of Motu they had acquired in Port Moresby, and the language was also carried to outlying areas by released prisoners (who themselves sometimes subsequently became policemen). This language spread throughout Papua as a result of the establishment of colonial authority through the police force. This is why this pidgin was initially named 'Police Motu' (instead of Pidgin Motu), a name that was shed, however, due to its colonialist connotations (see below).

Hiri Motu is currently used as a lingua franca throughout much of the former British territory of Papua, in the southern part of what is now Papua New Guinea. This is an area of immense linguistic diversity, with about 200 languages—some Austronesian, but many also highly varied non-Austronesian languages—spoken by several hundred thousand people. However, the English-derived pidgin language known as Tok Pisin is increasingly becoming the main lingua franca in areas surrounding the rapidly growing capital, Port Moresby, which is located within the traditional Hiri Motu area; it is also gaining speakers in some of the rural areas where it was not used formerly.

It is difficult to say at this stage what sort of future there is for Hiri Motu. Although it still has well over 100,000 speakers, competition with Tok Pisin is likely to slow its growth and possibly even cause its numbers to shrink. However, there has also been a degree of active resistance to the spread of Tok Pisin by some, who regard the English-based pidgin as a variety of 'broken English' imposed by the former colonial overlords. Hiri Motu, in contrast, has been seen as a genuinely 'Papuan' language, given its clearly non-English origins. This kind of political sentiment toward Hiri Motu probably peaked in the late 1970s around the

time that Papua New Guinea gained its independence from Australia (1975) with the development of the Papua Besena movement, and there has been increasing acceptance among Papuans since then of a genuinely national role for Tok Pisin.

In this political context, the original name 'Police Motu' also became quite controversial, and this name is seldom used in Papua New Guinea today. Prior to independence, the original British colony of Papua had largely been neglected, with little attention given to socioeconomic development. This contrasted sharply with the northern colony of New Guinea, which had originally been under German control (although after World War I, Australia took over). A vigorous plantation economy was established in coastal areas, and it was in this context that English-based Tok Pisin developed and spread throughout that colony.

After World War II, Papua and New Guinea were united into a single administrative entity, which was finally granted independence from Australia in 1975. However, many Papuans were uneasy about the prospect of becoming part of an independent country because they felt that they would be economically and numerically dominated by their northern neighbors, whose primary lingua franca, as it happened, was an English-based pidgin.

Papuans could already see the spread of Tok Pisin into areas that had traditionally made exclusive use of Hiri Motu as a lingua franca, and some Papuan nationalists declared their separate independence (but without recognition from any other nation). Hiri Motu was felt to represent Papuan nationalist aspirations in a way that Tok Pisin never could. Because of assumed associations with the colonial past, a government committee in 1971 officially renamed the language from 'Police Motu' to 'Hiri Motu', based on the assumption—albeit an inaccurate one, as emerged from Dutton's subsequent research—that this was the language of precolonial *hiri* trading expeditions.

References

Amini, Brian, Percy Chatterton, Raka Igo, Ron Lean, Api Leka, and Frank Wood. 1976. The dictionary and grammar of Hiri Motu. Port Moresby (Papua New Guinea): The Papua New Guinea Government Office of Information.

Dutton, Tom. 1985. Police Motu iena sivarai. Port Moresby (Papua New Guinea): The University of Papua New Guinea Press.

Dutton, T.E., and C.L. Voorhoeve. 1974. Beginning Hiri Motu. Canberra: Pacific Linguistics.

TERRY CROWLEY

See also **Austronesian; Tok Pisin**

Historical Linguistics

Languages are constantly in the process of change. Spellings of English homophones such as *meet* and *meat* reflect vowel distinctions at an earlier stage in the history of our language, before the two vowel qualities merged. Changes take place on the level of sound structure, grammar, vocabulary, and meaning. *Historical linguistics* is the study of language change by comparison of a language at two (or more) points in time.

The field of historical linguistics developed significantly in the nineteenth century, after European scholars realized that there were systematic similarities between Sanskrit, the ancient language of India, and other well-known languages, such as classical Greek and Latin. These similarities indicated that a genetic relationship must exist; i.e. these three languages must stem from the same ancestor language.

It was already well known that the Romance languages descended from a common language, Latin. A systematic comparison of *cognates*, words in related languages known to have been derived from a single historical source, led to the observation that sound change is, in fact, systematic. For example, there is a set of words starting with *h* in Spanish, such as *hijo* 'son', *hacer* 'do/make', and *harina* 'flour', that have cognates in Portuguese that start with *f*: *filho* 'son', *fazer* 'do/make', and *farinha* 'flour'. Through comparison with the words in Latin (such as *filius* 'son'), for which written records existed, it was possible to write rules to account for the various changes that must have taken place. Written records served two purposes: in addition to confirming the validity of the methodology being developed by historical linguists, they gave important information regarding the direction of sound changes. In the above examples, the Spanish *h* developed from Latin *f* and not the contrary.

The same methodology was then applied to the comparison of other languages that were obviously closely related, even though no written records existed

for the original language, or proto-language. For example, the Germanic languages were compared and classified as to how closely they were related to each other. The same was done for the Slavic languages.

Comparison within and between such groupings led to a classification of nearly all the languages in Europe, and some beyond, as members of the Indo-European family, all descended from a hypothetical language referred to as Proto-Indo-European, for which there are no written records.

In English, as well as other Indo-European languages, there are written records of the language at various points in time, making possible a comparison of earlier stages of the language with the present-day forms. The use of written records requires interpretation of the symbols appearing in them, to discover the sound structure of a language at the time of writing. In the case of the Germanic languages, the earliest written records use a different type of script, called *runes*, which themselves changed over time.

Comparison of the sounds in cognate words from related languages led to the development of a technique for *phonological reconstruction*. Sounds are studied systematically in their various positions within a word to obtain a full picture of what the original sound system must have been like that would account for the individual differences in the various daughter languages. Returning to the examples of Spanish and Portuguese, it would have to be noted that not all words with initial *h* in Spanish begin with *f* in Portuguese; some begin with *h*, as in the cognate pair *hombre* (Spanish) and *homem* (Portuguese) 'man'. Likewise, not all words beginning with *f* in Portuguese begin with *h* in Spanish; some begin with *f*, as in *fuego* (Spanish) and *fogo* (Portuguese) 'fire'. All of these patterns have to be examined. Furthermore, the best results require that attention be given to all descendant languages; in the case of Romance languages, this would include not only the languages with many speakers (Italian, French, Rumanian, Portuguese, and Spanish) but also those with fewer speakers (Catalan, Provençal, Rheto-Romance, and Sardinian). The precedent for the rigorous detail required in the comparative method was set by Jacob Grimm, in his work relating the German consonant system to the systems of other Indo-European languages.

Another method used by historical linguists is *internal reconstruction*. This involves the observation of patterns within an individual language to arrive at hypotheses regarding an earlier stage of that language, without taking into account information from outside languages. For example, the prefix *n-* ('not') in the Wayampi language has a variant *na-,* which occurs when the prefix attaches to a consonant. However, there are restricted cases in which the *na-* variant is used before a vowel, as in *orɨ* 'he is happy', *na-orɨ-i* 'he is not happy'. Through internal reconstruction, we can account for these exceptions by hypothesizing that there must have originally been a consonant (indicated by a capital C) between the prefix and the following vowel (**na-Corɨ-i*).

The study of the loss or retention of specific vocabulary items is another method used in historical linguistics, often called *lexicostatistics*. This can be used to study the degree of change at two stages of the same language or to study the degree of relatedness of two different languages. Usually, lists of basic core vocabulary are used for this type of study.

Dialect geography is another method used by historical linguists. Details of the pronounciation of selected words are plotted on maps throughout the area where a given language is spoken to determine the details and geographical boundaries of different dialects, such as Northern and Southern American English.

When different languages come into contact with each other, some degree of *lexical borrowing* inevitably takes place. The influence of French on English after the Norman invasion of the British Isles in 1066 was a major factor affecting English vocabulary, with words of French origin, such as *arise,* joining the already existing combinations of a verb and an adverb, such as *get up*.

One principle that motivates sound change is simplification. This may involve the weakening of more complex sounds to less complex counterparts, such as a change from the affricate tʃ to a simple fricative ʃ or s. Loss of consonants at the end of syllables or words is another example of simplification. On the grammatical level, simplification may involve the elimination of certain grammatical inflections of nouns. But simplification on one level may actually increase the complexity of a language on another level and trigger additional change.

One way by which simplification is achieved, particularly on the level of word structure, is through *analogy*. The extension of use of the plural suffix *-s* in English to words that originally had other plural endings has taken place by analogy. For example, the word *hippopotamus,* which came from Latin, originally took the plural form *hippopotami*. However, *hippopotamuses,* based on the main pluralizing pattern in English, is now the more commonly used form.

Analogy is also used in the creation of new words based on an already existing pattern in the language. The development of terms such as *chairperson* took place by analogy with the words they were replacing, in this case *chairman*.

Languages may change with respect to their sound structure or grammar. The meanings of words can

change, and old vocabulary items may be lost or new ones may be added.

The change of word-initial *f* to *h* in Spanish is one example of a sound change. Another change in this language is from *o* to *ue* in all words in which *o* had intonational stress (Latin *novem* 'nine', Spanish *nueve*). This change is *conditioned*; it occurs under specific conditions, i.e. only in particular phonetic environments. If all occurrences of *o* had changed to *ue,* without reference to any phonetic environment, the change would be *unconditioned*.

One type of grammatical change consists of the change or loss of affixes. Latin had a set of case suffixes that occurred on nouns to distinguish between subject, direct object, and indirect object, as in the word for 'daughter': *fili-a, fili-am,* and *fili-ae,* respectively. In Portuguese, *filha* is used without such case distinctions; i.e. Portuguese has lost the Latin case system. Other grammatical changes may affect pronouns, as in the elimination of *thee* and *thou* from most present-day dialects of English. The relative word order of the subject, object, and verb may also undergo modifications over time.

Words may change their meaning or develop secondary meanings. For example, the word *deer,* referring to a specific animal, developed from the more generic Old English word *dēor* 'beast'. In recent years, the word *mouse* has acquired an additional meaning, referring to a computer accessory.

The acquisition of new vocabulary may take place through innovation (such as *byte*)*,* blending of already-existing vocabulary (*brunch*, from *breakfast* and *lunch*), or the use of derivational processes existing in the language (*computer* from *compute*). It may also take place through *lexical borrowing* as the result of language contact. Some examples of such borrowing into English include *canoe* (borrowed from Carib), *jaguar* (from Tupinambá), *safari* (from Swahili), and *smorgasbord* (from Swedish).

As changes take place in different geographical regions where a single language is spoken, different dialects develop. If speakers of different dialects are isolated from each other over long periods of time, the dialects may undergo so many changes that speakers from the different groups no longer understand each other. At this point in time, what were originally different dialects of the same language have become different languages, which are members of a linguistic family or possibly a subgroup within the family. For example, the Scandinavian languages (Danish, Icelandic, Norwegian, and Swedish) make up a subgroup within the Germanic language grouping. They are much more closely related to each other than they are to English, German, or Dutch, having descended from a common ancestor language referred to as Old Norse.

A detailed comparison of individual languages showing a high number of cognate words provides the information necessary for a classification and subclassification within a language family. A diagram, similar to a family tree, can then be used to show the internal *classification* of the family. This type of detailed classification was first done for Indo-European by the nineteenth-century historical linguists.

The methods of historical linguistics are now being applied to the study of many indigenous languages and language families. It is possible to recognize languages that are similar to each other, by identifying a significant number of cognates (words coming from a single source). Through a systematic comparison of the similarities and differences in the sound structure of cognates, it is possible to reconstruct what the sound system of the common ancestor, or 'proto-language', must have been like to account for the present form in the various languages. Protoforms are preceded with an asterisk to indicate that they are hypothetical. For example, in proto-Tupi-Guarani, of lowland South America, one of the reconstructed consonants is **ts,* which has reflexes of *ts, s, h*, and *0* (*zero*) in individual languages of the family.

Phonological reconstructions have been done for several indigenous language families of the Americas, as well as in other parts of the world. The quality of a reconstruction depends on the availability of adequate and accurate data of individual languages within a given family. As more data become available, the reconstructions are revised as necessary, and subclassification is possible.

The word and sentence structures of some indigenous language families, such as the Tupi-Guarani family in South America, have also been reconstructed. Grammatical reconstructions require much more data than do reconstructions of sound and are sometimes done in stages, as an increasing number of grammatical descriptions of individual languages are made available. To give an example, the set of pronouns and personal prefixes was first reconstructed for proto-Tupi-Guarani without reference to the particular grammatical contexts in which they were used. Additional data were necessary before a more complete description could be made.

Summarizing, the methods of historical linguistics, which were developed and tested in the study of Indo-European languages, are now being applied to the study of languages in other parts of the world.

References

Bynon, Theodora. 1977. Historical linguistics. London: Cambridge University Press.

Campbell, Lyle. 1999. Historical linguistics: an introduction. Cambridge, MA: MIT Press.

Jeffers, Robert J., and Ilse Lehiste. 1979. Principles and methods for historical linguistics. Cambridge, MA: MIT Press.

Jensen, Cheryl. 1999. Tupi-Guarani. The Amazonian languages, ed. by R.M.W. Dixon and Alexandra Y. Aikhenvald. Cambridge: Cambridge University Press.

Lehmann, Winfred P. 1993. Historical linguistics: an introduction. London and New York: Routledge.

McCrum, Robert, William Cran, and Robert MacNeil. 1993. The story of English. New York: Penguin USA.

Trask, R.L. 2000. Dictionary of historical and comparative linguistics. Chicago: Fitzroy Dearborn Publishers.

CHERYL JENSEN

History of Linguistics: Overview

The Greco-Roman World

The European tradition of linguistics begins in Classical Greece with the development of an active philosophical discourse on linguistic topics. Several of Plato's dialogs touch upon linguistic issues, with one, the *Cratylus*, devoted exclusively to the subject of the conventionality of language. The issue is discussed along the lines of a genealogy of language, with one side supporting a 'natural' origin for human communication, and another a 'conventional' basis. In the same context, onomatopoeia (sound symbolism) and the internal relation between names and their referents are discussed, with mixed conclusions. From the outset of the Greek tradition, questions about language concentrated entirely on the Greek language and were considered within the terms of the nature–convention debate and the somewhat related regularity–irregularity (analogy–anomaly) controversy. Aristotle also discussed linguistic issues in his treatises on logic and rhetoric—without dedicating any of them exclusively to the subject—with a discernible overall tendency to subscribe to conventionalist doctrine.

After Aristotle, the Stoics, with their founder Zeno (c. 315 BCE), formalized the dichotomy between form and meaning by distinguishing in language the 'signified' and the 'signifier'. They gave separate treatment to phonetics, grammar, and etymology, devoting considerable attention to the last. The Stoics accorded grammar-independent recognition within philosophical studies; it is, however, difficult to reconstruct with certainty the details of their linguistic theory.

The main aspects of linguistic study that received specific attention among early Greek scholars were etymology, phonetics, and grammar. Etymology was understood as the unfolding of words, through which their true meanings are made plain. In phonetics, some articulatory classifications of speech sounds were attempted, the syllable was introduced as a unit of description, and by the time of the Stoics there was a rudimentary understanding of speech as the effect of articulatory interference with the flow of air from the lungs. The descriptive framework for Greek phonetics was the Greek alphabet, and statements took the form of accounts of the pronunciation of the letters in it. Plato grouped together vowels in contrast to consonants and distinguished within consonants between continuants and stops, the latter being unpronounceable without an adjacent vowel sound. The Stoics distinguished three aspects of the *gramma*, thereby confirming its phonetic-orthographic unity: its phonetic value, its written shape, and the name by which it was designated. These three properties of letters continued to be distinguished throughout Antiquity and the Middle Ages, their later Latin names being *potestas*, *figura,* and *nomen*. The Greco-Roman world did its best and most influential work in the field of grammar. Dionysius Thrax's *Techne Grammatike* (most of it of debated authorship) and Apollonius Dyscolus's *Syntax* are the two most significant works in the field.

The Romans built largely on the Greek foundations. Varro—with his *De Lingua Latina*, of which only a part survives—is the first attested Latin writer on linguistic questions who is of importance. One of the main sources for the analogy–anomaly controversy, he may have misinterpreted it as a matter of permanent academic attack and counterattack, rather than as the more probable coexistence of opposing tendencies. Varro envisaged language developing from an original limited set of primal words, imposed on things so as to refer to them, and acting productively as the source of large numbers of other words through subsequent changes in letters, or in phonetic form, as civilization advanced and human life became richer. Apart from his innovative distinction between derivational and inflectional word formation, Varro also showed originality in his proposed classification of inflected words in Latin, by setting up a quadripartite system of four inflectionally contrasted classes.

Priscian's voluminous grammar (c. 500 CE) may be taken as representative of late Latin grammarians. The Greek technical terms are given fixed translations with the nearest available Latin word. Priscian draws heavily on Apollonius and his son Herodian, acknowledging both in his introductory paragraphs. He organized the description of the forms of nouns and verbs and of the other inflected words, by setting up canonical or basic forms; from these, he proceeded to the other forms by a series of letter changes, the letter being for him, as for the rest of Western Antiquity, the minimal unit of both orthography and pronunciation.

The Middle Ages

Continuing from the Classical tradition, the Middle Ages had a distinct focus on grammar as a foundation of scholarship, both as itself a liberal art and as a prerequisite for reading and writing Latin correctly. Linguistic studies, almost wholly pedagogical in their aims and largely derivative in their doctrine, were subordinate to theology, the Christian faith and Christian doctrine, and closely followed Donatus and Priscian. In the history of Medieval linguistic science, the most exciting period is from around 1100 to the end of the period (and more specifically between c. 1200 and c. 1350), when scholastic philosophy gave rise to a number of 'speculative grammars' or treatises *De modis significandi*. Scholasticism itself was the result of the integration of Aristotelian philosophy into Catholic theology. The mere description of Latin was considered inadequate, and attention was paid to the mental representations to which speech relates. The main distinction was between *significatio* (the meaning of a word) and *suppositio* (its point of reference in the real world). This basic distinction appeared repeatedly in different forms and with different interpretations, in oppositions such as meaning–reference, denotation–connotation, and intension–extension. In the modistic system, entities possess various properties (*modi essendi*). The mind apprehends these in certain modes of understanding (*modi intelligendi*). In language, the mind confers meaning on successions of sounds (*modi significandi*), which thus become words. The demand that grammatical description should be integrated into philosophical theory and represent universal principles of cognition brought about a great change in linguistic studies. Unlike all previous grammatical doctrines, the *modistae* were theory- rather than data-oriented.

The Renaissance and Modern Times

The adherence of linguistic analysis to the Greco-Roman archetype became less strict in the fifteenth and sixteenth centuries. Grammatical studies of Hebrew, Arabic, and modern European languages register a shift in the linguistic interests of Europeans and trigger creative thinking on language structure and written representation. With the invention of printing and the emergence of a need for uniform orthography within each language, pronunciation issues gain greater attention.

By the end of the sixteenth and the beginning of the seventeenth centuries, the indigenous languages of the Americas had captured the imagination of Europeans. Colonization of the Far East meant that the Chinese languages and their writing system became known to Europe and were largely admired. With very different word structures those of European languages, Chinese and Native American languages served to stretch the linguistic experience of Europeans far beyond what the Classical and Medieval world had allowed them. At the same time, the Royal Society in Britain (established 1660) and the Académie Française (established 1635) in France were both concerned with linguistic research and with literary and linguistic standards.

In the intellectual movements of the sixteenth and seventeenth centuries, both rationalist and empiricist influences may be seen at work. The notion of a universal thought possessed by mankind, basically independent of any particular language and therefore expressible in a universal language, such as that of John Wilkins, was a rationalist conception. Rationalism is also responsible for the production of philosophical grammars—especially those associated with the French Port Royal schools—which succeeded the Medieval scholastic grammars. They attempted to reveal the unity of grammar underlying the separate grammars of different languages in their role of communicating thought, itself comprising perception, judgment, and reasoning.

As part of English empiricism, work was carried out on phonetic questions under the titles of 'orthography' (right spelling) and 'orthoepy' (right pronouncing) during the same period (the term 'phonetics' is first recorded in the nineteenth century). Work by William Holder (1669) and Francis Lodwick (1686) revolutionized articulatory phonetics, and John Wilkins included in his *Essay towards a real character and a philosophical language* (1669) a sound chart that can be compared with early editions of the International Phonetic Alphabet.

In the eighteenth century, linguistic issues again became central for philosophy as part of its search for the traits of primitive society. Both Etienne Condillac and Jean-Jacques Rousseau considered that abstract vocabulary and grammatical complexity developed from an earlier individual concrete vocabulary with very few grammatical distinctions or constraints. They also regarded reliance on tonal contrasts in the

manner of Chinese as a survival of a primitive feature and considered poetry to have sprung from chanting as the earliest form of language.

In 1772, Johann Herder supported a monogenetic (common origin) theory of all languages, as of all cultures. His theory suffered from the restricted time perspective with which the eighteenth century viewed man's existence on earth, with a consequent attempt to see enduring characteristics of the early stages of language in alleged 'primitive languages' still existing in the present. Work by James Harris (1751) and Horne Tooke (1786) demonstrate the vivid interest of the time in the origin of language.

Wilhelm von Humboldt—exceptionally for his time—did not concentrate predominantly on history. He did not distinguish sharply between systemic and historical perspectives and followed Herder in asserting the individuality of each different language as a peculiar property of the nation or the group that speaks it, despite believing in the universality of speech. Humboldt is best known in linguistics for popularizing a tripartite language typology (isolating, agglutinative, and flexional) according to the predominant structure of the word as a grammatical unit.

Nineteenth Century: Comparative and Historical Linguistics

In the last quarter of the eighteenth century, linguistic science received a very productive stimulus from India. In 1786, Sir William Jones showed the historical kinship of Sanskrit, the classical language of India, with Latin, Greek, and the Germanic languages. India was not just the source of linguistic evidence but also offered an alternative linguistic tradition, dating back to the second half of the first millennium BCE, with outstanding achievements in general linguistic theory and semantics, phonetics and phonology, and descriptive grammar.

Early comparativists concentrated on Sanskrit and its relations with the other Indo-European languages (hence the term *vergleichende Grammatik* 'comparative grammar'). Rasmus Rask (1787–1832) was the first to bring order into etymological relationships by setting out systematic comparisons of word forms, matching a sound in one language with a sound in another, with such matches exemplified in a number of different words. The correspondences now known as Grimm's Law were in fact first discussed by Rask.

The historicism and nationalism of the time characterized the work of Franz Bopp and Jacob Grimm. Bopp saw as the main purpose of his *Conjugation system* the reconstruction of the original grammatical structure of the language whose gradual disintegration had produced the attested languages of the Indo-European family. Both the use of comparison as the clue to earlier history and the conception of change as degeneration from primitive integrity were common property of the scientific thought of the time. Relations between the parent language and the known Indo-European languages were set out according to the *Stammbaumtheorie* or genealogical tree model. August Schleicher, with his biological approach to language and his quest for the *Ursprache* (the original language), initiated the practice of distinguishing reconstructed forms with an asterisk (whence the later term 'starred forms'). He assigned Sanskrit a place like any other language, in the 'Arian' (Indo-Iranian) group and expressed his belief that the three current language types (isolating, agglutinating, and inflectional) represent historical stages in the growth of languages to their highest point of organization.

The Neogrammarians also strove to ground their work on comparative–historical linguistics within the natural sciences, preferring physics over biology as their model. Hermann Osthoff and Karl Brugmann set out the principles of the movement in a programmatic article (1878): all sound changes, as mechanical processes, take place according to laws that admit no exceptions (*ausnahmslose Lautgesetze*) within the same dialect and within a given period of time; the same sound in the same environment will always develop in the same way; analogical creations and reformations of specific words as lexical or grammatical entities are a universal component of linguistic change at all periods of history and prehistory. The influence of the Neogrammarians in linguistic practice has been overwhelming. It is true, however, that they were largely drawing out what had been implicit in the very practice of the subject, and distinguishing it from unnecessary and fallacious assumptions.

The critique of the Neogrammarian position came from a branch of linguistics that they had been at pains to encourage, dialectology. The temporal and geographical limits of a dialect are difficult to determine, especially when one scrutinizes the language closely. The most important opponents were Hermann Schuchardt, who included in his works an article 'On Sound Laws: Against the Neogrammarians,' and Jules Gilliéron, who put together the linguistic atlas of France.

Twentieth Century

The key figure in the change from nineteenth- to twentieth-century attitudes was the Swiss linguist Ferdinand de Saussure, who after studying in Leipzig with members of the Neogrammarian school made his first significant contribution in Indo-European comparative linguistics. Saussure's doctrine on linguistics in the early twentieth century was recorded in the *Cours de*

linguistique générale (1916; *Course in General Linguistics*), a reconstruction of his lectures from notes taken by his students, published posthumously. Saussure's contribution lies in three major areas: (1) He formalized the two fundamental dimensions of linguistic study: 'synchronic', in which languages are treated as self-contained systems of communication at any particular time; and 'diachronic', in which the changes to which languages are subject in the course of time are treated historically. (2) He distinguished the linguistic competence of the speaker as a member of a speech community from the actual phenomena or data of linguistics (utterances), as, respectively, *langue* and *parole*. (3) He showed that any *langue* must be envisaged and described synchronically as a system of interrelated elements, lexical, grammatical, and phonological, and not as an aggregate of self-sufficient entities (which he compared to a mere nomenclature). The greatest impact of structuralism was in the domain of phonology (the study of sound systems), where Nikolaj Trubetzkoy and Roman Jakobson developed the theory of 'distinctive features'.

'Structuralism' was widely preached in America and was epitomized in the historical figure of Leonard Bloomfield. American structuralism, however, differed from its European ancestor in its insistence on a data-oriented approach which originated from the urgent need to collect data on the disappearing Native American languages. This persistence remarkably limited the theoretical perspectives of American structuralism and eventually brought about its downfall.

The reaction came in the middle of the century and resulted in a profound change in the outlook of theoretical linguists worldwide. Noam Chomsky's *Syntactic structures* (1957) marks the first public appearance of a change in outlook on the study of language and the scientific status of linguistics. His theory, 'generative linguistics', is still developing and has become the dominant theory in linguistics at the end of the twentieth century. It involves the notion of a linguistic 'deep structure', whence utterances are produced through a series of 'transformations'. Transformations became less crucial in subsequent versions of the theory and the lexicon as a source attained greater importance.

In the last quarter of the twentieth century, linguistics has gained still greater status as a pioneering academic discipline. An increased diversity—stemming from the emergence of branches such as computational and applied linguistics and psycholinguistics alongside the established subdisciplines—is the most prominent and promising feature of the discipline at the turn of the millennium.

References

Baratin, Marc, and Françoise, Desbordes. 1981. L'analyse linguistique dans l'antiquité classique. Paris: Klincksieck.

Bursill-Hall, Geoffrey Leslie. 1971. Speculative grammar in the Middle Ages: the doctrine of Partes orationis of the Modistae. The Hague: Mouton.

Chomsky, Noam. 1957. Syntactic structures. The Hague: Mouton.

Eco, Umberto. 1995. The search for the perfect language. Oxford and Cambridge, MA: Blackwell.

Humboldt, Wilhelm von. 1988. Über die Verschidenheit des menschlichen Sprachbaues; as On language: the diversity of human language-structure and its influence on the mental development of mankind, translated by Peter Heath, with an introduction by Hans Aarsleff. Cambridge: Cambridge University Press.

Law, Vivien. 1997. *Grammar and grammarians in the early middle ages*. London and New York: Longman.

Lepschy, Giulio (ed.) 1994–1998. History of linguistics, Vol I: The eastern traditions of linguistics, Volume II: Classical and medieval linguistics, Volume III: Renaissance and early modern linguistics, Volume IV: Nineteenth-century linguistics. London and New York: Longman.

Matthews, Peter Hugoe. 1993. Grammatical theory in the United States from Bloomfield to Chomsky. Cambridge: Cambridge University Press.

Pedersen, Holger, and John Webster Spargo. 1962. The discovery of language, linguistic science in the nineteenth century, 3rd edition. Bloomington: Indiana University Press.

Robins, Robert Henry. 1967. A short history of linguistics. London and New York: Longman; 4th edition, 1997.

Saussure, Ferdinand de. 1983. Cours de linguistique générale. Paris: Payot, 1916; as Course in general linguistics, translated and annotated by Roy Harris. London: Duckworth.

Ziolkowski, Jan (ed.) 1990. On philology, University Park and London: Pennsylvania State University Press.

CHRISTOS G. NIFADOPOULOS

Hittite

Hittite belongs to the Anatolian branch of the Indo-European family. It was spoken in inner Anatolia during the second millennium BCE, with its written sources ranging from approximately 1650 to 1180 BCE. It is the oldest attested Indo-European language. Other Anatolian languages include Palaic and

Cuneiform Luwian (also attested in the second millennium BCE), Hieroglyphic Luwian, Lycian, Milian, Carian, and Lydian (attested in the first millennium BCE); all these languages are relatively poorly known.

Knowledge of Hittite is a recent achievement: although some Hittite texts were found much earlier, they were deciphered starting from 1916 only. Hittite is documented by several thousands of clay tablets, mostly found in the archives of the Hittite capital, Hattusa, near the modern village of Bogazkahle, about 210 km southeast of Ankara. The Hittites wrote their language in cuneiform, a syllabic system constituted by signs for CV, VC, V, and, to a limited extent, CVC syllables, also used for numerous other languages in the Ancient Near East. The cuneiform syllabary also contains a number of ideograms, conventionally rendered in transliteration with the Sumeric equivalent (in capital letters, see e.g. (1) and (2) below); ideograms were the key for the decipherment of Hittite.

Grammatical sketch

Phonology

Partly for problems connected with syllabic writing, partly on account of as yet poorly understood scribal habits, our understanding of Hittite phonology is not completely clear. Among vowels, four separate phonemes certainly existed: /a/, /e/, /i/, /u/; the existence of a phonemic /o/ is still discussed. Vowel length was probably distinctive, as shown by pairs as *uddar*, 'word', *uddār*, 'words'.

Consonants include four series of stops, bilabial, dental, velar, and labiovelar, all with a distinction between voiced and a voiceless (or lax and tense), which was apparently neutralized in the word-initial and word-final positions. Simple vs. double writing of stops and of the fricative -*h*- renders the voiced ~ voiceless contrast. The fricative /s/, liquids and nasals could be distinctively simple or geminated word-internally (Table 1).

Morphology

Hittite is an inflectional language of the fusional type. Inflectional categories of nouns include case (nominative [NOM], accusative [ACC], genitive [GEN], dative/locative [D/L], directive, ablative, instrumental, vocative) and number (singular and plural); nouns belong to two genders: common and neuter. The directive and the instrumental case are mostly limited to the ancient language (Old Hittite); later (Middle and New Hittite), they merge with the dative/locative and the ablative, respectively. Directive, ablative, and instrumental only had endings for the singular; in Old Hittite, they could never occur with animate nouns (the animate vs. inanimate distinction cross-cuts grammatical gender, inanimate nouns belonging partly to the neuter, partly to the common gender).

An original feature of Anatolian is the use of derivational suffixes to serve syntax, rather than to enrich the lexicon, a function normally assigned to inflectional morphology. Thus, in Cuneiform Luwian, there is no inflectional genitive; nouns used as modifiers of other nouns take a derivational suffix, which allows them to behave as adjectives, and agree with their heads. In Hittite, neuter nouns cannot function as subjects of transitive verbs, but must obligatorily be transposed to the common gender by means of the suffix -*ant*-: after suffixation, -*ant*- derivates take the nominative ending of the common gender and can be made the subject of transitive verbs (e.g. *tuppi*, 'clay tablet', neuter, *tuppiyanza*, same meaning, common gender).

Adjectives and pronouns inflect for case, number, and gender; attributive adjectives usually agree with their head. There is no specific morphology for comparison. Personal pronouns have stressed and unstressed forms; an unstressed nominative only exists for third-person pronouns. Besides, Old Hittite also has enclitic (i.e. unstressed, hosted by the preceding word) possessives inflected to agree with their head, which also functions as a phonological host: *attas=mas*, 'of my father' (lit.: father-GEN my-GEN; the sign = indicates enclisis); after Old Hittite, genitive forms of stressed pronouns are used instead of possessives: *ammel attas*, '(of) my father' (lit.: I-GEN, father-NOM (or GEN)).

Inflectional categories of verbs include tense (present [PRES]/future and preterit [PRET]; there is a periphrastic perfect with the auxiliaries *har(k)*-, 'have', and *es*-, 'be'), mood (indicative and imperative; other types of modality are expressed by the particle *man* and the indicative), diathesis (active and medio-passive), person, and number. Nonfinite verb forms are the infinitive, the participle, and the -*wan*- supine, only used in inchoative constructions; verbal nouns are built with the suffixes -*atar* or -*uwar/-mar: iya*-, 'to make', *iyawar*, 'the act of making'. Verbs are divided into two inflectional classes, called -*mi* and -*hi* conjugation from the ending of the first person singular. Some other verbal categories are expressed derivationally. Causatives are made with the suffix -*nu*-: *huinu*-, 'to let escape', from *huwai*-, 'to run', *huisnu*-, 'to rescue', 'to

TABLE 1 Hittite Consonants

Voiceless stops	p	t		k	kʷ
Voiced stops	b	d		g	gʷ
Voiceless fricatives		s			h
Voiced fricative					ɦ
Affricate		ts			
Nasals	m	n			
Lateral		l			
Vibrant		r			
Glides	w		j		

give life', from *huis-*, 'to live', 'to be alive' (this suffix can also derive verbs from adjectives: *daluganu-*, 'to lengthen', from *daluki-*, 'long'). The suffix *-sk-* is used for iteratives: *pai-*, 'to give', *pesk-*, 'to give habitually'.

Derivation is widely used for creating new words; derivational affixes are by the most part suffixes. (There are two verbal prefixes, *pe-* and *u-*, which mostly occur with motion verbs and express opposite deixes; verbal reduplication is not productive: often, reduplicated verbs have the same meaning as their nonreduplicated counterparts.) Composition is comparatively infrequent.

Syntax

Hittite has nominative–accusative alignment for common gender nouns. As remarked in the preceding section, neuter nouns must be transposed to common gender in order to function as the subject of transitive verbs. Since this kind of derivation has a systematic character, the derivational affix together with the common gender nominative ending functions as an ergative case. A further distinction between transitive and intransitive subjects, only involving third person, is that transitive subjects can be omitted if they are recoverable from the context, whereas intransitive subjects cannot.

Hittite has a rigid Object-Verb order, with a verb-initial variant sometimes found under specific pragmatic conditions (see e.g. (1a)). Modifiers (attributive adjectives, genitives, relative clauses) regularly precede their head, but certain determiners follow. Adpositions follow their complements (postpositions), and inflected auxiliaries follow nonfinite forms of main verbs. All enclitics, except for possessives, the focalizer *-pat*, and conjunctions used in noun phrase coordination, follow Wackernagel's Law very strictly; consequently, they are placed after the first accented word in the sentence (P2). Sentence connectives (CONN) occur in virtually every sentence after the Old Hittite period; often, it seems that they are introduced for the only purpose of hosting enclitics. The following example serves as an illustration of Hittite sentence structure:

(1 a) *sallanun=* *war= an* *kuit*
promote:1SG.PRET PTC 3SG.ACC because
ammuk
1SG.NOM

(b) *nu=* *war=an* *huwappi* DI-*esni*
CONN PTC 3SG.ACC bad:D/L tribunal:D/L
huwappi DINGIR^LIM-*ni* UL *para*
bad:D/L god:D/L NEG PREV
UL kuwapikki *tarnahhun*
never hadle:1SG.PRET

(c) *kinuna=ya=* *war=an* *karapmi*
now and PTC 3SG.ACC take:1SG.PRES

(d) *nu=* *war=an* *ANA* ^DUTU
CONN PTC 3SG.ACC to sungod
^URUTÚL-*na* *AŠŠUM*^LÚ SANGA-*UTTIM*
Arinna for priesthood
tittanumi
install:1SG.PRES
'(a) because I promoted him, (b) I never handled him over to a bad tribunal or to a bad god; (c) and now I will take him (d) and make him priest for the sun goddess of Arinna', *StBoT* 24, IV 11–15.

Note: NEG = negation; PREV = preverb; PTC = particle; SG = singular

Hypotaxis is relatively poor, the most frequent type of subordinate clause being relative clauses. Relative clauses are corelative in the great majority of occurrences: the relative clause, mostly preposed, contains its head, which may or may not be referred to by a demonstrative in the main clause. An example is given below:

(2) ^GIŠTUKUL= *ma* *kuin*
weapon PTC REL.ACC
apiya harkun *n= an*
there have-1SG.PRET CONN 3SG.ACC
halissiyanun
inlay-1SG.PRET
'the weapon that I had there I had inlaid', *StBoT* 17, obv. 46.

Note: REL = relative

Circumstantial clauses include temporal, causal, and conditional clauses; they are mostly preposed. Complement clauses are virtually nonexistent; only in New Hittite a limited number of postposed complement clauses occurs, introduced by the conjunction *kuit* (nominative/accusative neuter singular of the relative pronoun, already used in causal clauses with the meaning 'because', cf. e.g. (1a)).

Hittite is extremely rich in enclitic particles. Among P2 enclitics, one finds adversative connectives (as *-ma* in e.g. (2)), the direct speech particle *-wa(r)-* (cf. e.g. (1)), the reflexive particle *-za*, and the so-called local particles, most likely cognate to the Indo-European preverbs (the modal particle *man* can be enclitic, and occur P2, or sentence initial and host the other enclitics; enclitic personal pronouns also occur in P2, cf. the form *-an* in the examples).

Hittite and Indo-European

When Hittite was deciphered, it immediately became clear that it did not fit into the traditional reconstruction of Proto-Indo-European (PIE), mostly based on Greek and Sanskrit. The absence of a number of grammatical categories traditionally reconstructed for PIE (optative,

aorist, dual, and the three-gender system, among others) has given rise to different theories. The most well-known is the so-called 'Indo-Hittite hypothesis', put forward by Sturtevant in 1933. According to Sturtevant, Anatolian must be considered a separate branch of a language family whose other branch is constituted by all other Indo-European languages: the nonoccurrence of certain features in Anatolian can then be explained by separate developments of the Indo-European branch. Since the publication of Sturtevant's book, the understanding of Hittite has become deeper, partly thanks to the availability of newly found material; a better insight into the position of Hittite among the other Indo-European languages has been gained after cuneiform paleography has allowed a precise dating of the Hittite tablets. Many other theories have been worked out, among which was the so-called *Schwundhypothese* ('loss hypothesis'), according to which Hittite once had all the inflectional categories known from the other ancient Indo-European languages, but lost them, as a result of intensive contact with non-Indo-European languages in prehistoric times.

However, the discovery of Hittite did not only come to raise problems for the traditional reconstruction of PIE: many important findings provided evidence for theories that could scarcely be demonstrated before. The most striking confirmation was given by the reflexes of the PIE laryngeals, reconstructed on the basis of Saussure's *coefficients sonantiques* (Saussure 1878). Indirect evidence for these phonemes had been found in the other languages, but only Hittite preserves two of them as phonemes, also providing evidence for their consonantic character. It is nowadays commonly thought that the PIE phoneme inventory included three laryngeals, conventionally written as $*/h_1/$, $*/h_2/$, and $*/h_3/$. While in the other Indo-European languages laryngeals only affected the following or preceding vowel, in Hittite reflexes of $*/h_2/$ and $*/h_3/$ are found as *-h-*. Reflexes of laryngeals are found in the following examples: Hitt. *harki* 'white' (Latin *argentum*); Hitt. *hastai* 'bones' (Ancient Greek *osteon*); Hitt. *hulana* 'wool' (Sanskrit *urna*); and Hitt. *huis-* 'to live' (Latin *uiuere*). After a vowel, $*/h_2/$ is also preserved as /ḫ/: Hitt. *newahh-* 'renew', < $*neweh_2$- (Latin *nouāre*).

Another important confirmation came from the placement of enclitics in P2: Wackernagel (1892) had formulated his law on word order mostly on the basis of Ancient Greek and Sanskrit, where enclitics often occur in P2, but much less frequently than in Hittite.

The Hittite verb also preserves a number of archaic features, notably the athematic (*-mi*) conjugation; the *-hi* conjugation is a Hittite innovation, related to both the PIE perfect and the PIE middle.

Among indisputable innovations, one can list the emergence of split ergative alignment, and the constraints on omission of third-person subjects.

References

Chicago Hittite dictionary, Chicago, 1980.

Friedrich, Johannes. 1952. Hethitisches Wörtebuch. Heidelberg: Winter.

Friedrich Johannes. 1960. Hethitisches Elementarbuch. Heidelberg: Winter.

Held, Warren. 1957. The Hittite relative sentence. Baltimore: LSA (Language Dissertations 55).

Hrozný, Friedrich. 1917. Die Sprache der Hethitier. Leipzig: Hinrichs.

Keilschrifttexte aus Bogazköy. Leipzig, 1916–23; Berlin, 1954.

Keilschrifturkunden aus Bogazköy. Berlin, 1921.

Kronasser, Heinz. 1966. Etymologie der hethitischen Sprache. Wiesbaden: Harrassowitz.

Luraghi, Silvia. 1990. Old Hittite sentence structure. London-New York: Routledge.

Luraghi, Silvia. 1997. Hittite. Munich–Newcastle: Lincom Europa.

Melchert, H. Craig. 1994. Anatolian historical phonology. Amsterdam: Rodopi.

Saussure, Ferdinand de. 1878. Mémoire sur le système primitif des voyelles dans les langues indo-européennes. Leipzig: Teubner.

Studien zu den Bogazköy Texten, Wiesbaden: Harrassowitz, 1965.

Sturtevant, Edgard A. 1933. A comparative grammar of the Hittite language. Philadelphia: Linguistic Society of America.

Wackernagel, Jacob. 1892. Über ein Gesetz der indogermanischen Wortstellung. Indogermanische Forschungen 1. 333–436.

SILVIA LURAGHI

Hjelmslev, Louis

Among post-Saussurrean European structuralists, Louis Hjelmslev is the one who took the most extreme positions with respect to the separation between form and matter, and system and use. The son of a mathematician, he always denied any influence from his father for his interest in abstract structures: in fact, the son's attempt at

creating a system independent of experience was criticized even by the father, who studied the relation between geometry and practical experience.

In spite of his denial of the value of experience, Hjelmslev had a deep knowledge of numerous languages, in some of which he was also fluent. According to an often quoted anecdote, he used to pay his parent's maid in order to give her Italian lessons. Later on, his interest for language diversity led him to write some insightful pages on the purposes of language typology in the last chapter of his book *Sproget* (1963), and to explore the grammar of a number of exotic languages (for example in his 'La catégorie des cas', published in 1935 and reprinted in the 'Essais linguistiques').

Hjelmslev illustrated Saussure's claim that language is formed by working out a model of language based on the co-occurrence of two elements: expression and content. Each element consists of form and substance, e.g. sound is the substance of utterances, but their form is articulated in phonemes. Although for Saussure, neither form nor substance is independent of the other, and it is the conjunction of two forms, form of expression and form of content, which generates a linguistic sign, Hjelmslev's treatment clearly implies a logical priority of form over substance, contrary to Saussure's view. Note that the interdependence of form and substance reveals an Aristotelian view of matter: as Hjelmslev explained in his *Prolegomena to a theory of language*, substance can be studied only when molded by form, and beyond form it consists of an unanalyzable 'purport'.

On account of his interest being centered only on structured, Hjelmslev took a completely different path from the one taken by Saussure's own pupils, as Bailly and Séchehay, who pursued a 'linguistique de la parole', thus aiming at the description of language usage. He was also critical of the views held on phonology by members of the Prague Circle, because they paid too much attention to the phonetic substance of phonemes. His major emphasis was on the nature of a possible theory of language, which should be arbitrary and adequate, and also aim at simplicity (in the case that two available explanations yield the same results, the one that requires less complex description must be preferred).

As is often the case in the history of linguistics, Hjelmslev's thought found a fertile soil in his cultural environment, and he could develop his ideas in close collaboration with a number of other scholars, who shared his beliefs about the strictly formal and synchronic nature of languages as systems. The outcome of scientific intercourse was the creation of the Linguistic Circle of Copenhagen in 1931, on the model of the Prague Linguistic Circle, which had been founded five years before. The first result of collaboration among the Danish scholars was presented at the international conference on phonetic sciences in London in 1935, where Hjelmslev, together with P. Lier and H. Uldall, illustrated a new phonological theory, called 'phonematics'. Shortly thereafter, in December 1935, Hjelmslev and Uldall presented a unified theory of phonology and grammar, called 'glossematics', which aimed at the study of mutual relations between phonemic and grammatical systems.

In spite of the ambitious program put forth by its founders, the theory of glossematics was never fully developed. In its final shape, glossematics should have developed a metalanguage, capable of describing all existing languages in algebraic terms. Since all linguistic elements must be described only in terms of the relations holding between one another, a typology of possible relations needed to be set up prior to the study of languages. In his *Prolegomena to a theory of language*, Hjelmslev describes the possible types of relations. Crucial to the understanding of his classification is the notion of 'function'. A function is a relation between linguistic entities; since they are considered only for the sake of functions, concrete linguistic entities are called 'functives'. In the case of (superordinate) functions holding between other functions, the latter can become functives. In his *Prolegomena to a theory of language*, Hjelmslev defines three types of function: (a) interdependence, where A implies B and B implies A; (b) determination, where A implies B, but B does not imply A; and (c) constellation, where A does not imply B and B does not imply A. It is important to stress that this use of the term function differs significantly from the way in which the same term was defined by the Prague Circle, which is certainly more familiar. One of the basic tenets of the Prague Circle was that language must be investigated in its nature of 'functional system', whereby 'functional' refers to the communicating function of language, with an implication of teleology. In the theory of glossematics, on the other hand, there is no space for teleology, and no intrinsic interest toward communication.

According to the promoters of glossematics, a second stage in the development of the theory should be based on the scrutiny of the largest possible number of languages, in order to assess what types of grammatical elements can occur in language. After the metalanguage had been worked out completely, it could be applied to the description of any given language.

The theoretical foundations of glossematics, which should have been published by Hjelmslev and Uldall, never saw the light because of disagreement between the two, and the only published part, the introduction, was the work of Uldall alone (Uldall, 1957). The only

thorough description of a language following the theory of glossematics is Togeby's *Structure immanente de la langue française*. Although Hjelmslev's own focus had been primarily on phonology and morphology, Togeby's book shows that glossematics was by no means limited to these parts of grammar, and takes the text as its starting point; it also offers a discussion of syntax, consisting of a description of subordinate clauses and VP and NP structure.

Biography

Louis Hjelmslev was born in Copenhagen on October 3, 1899, as Louis Trolle Petersen (the name Hjelmslev was added in 1903 by his father, after the village where he was born). From 1917 to 1923, he studied in Copenhagen (Master of Arts), where in 1920 he received an award for work on Oscan inscriptions. He pursued further studies in Prague (1923–1924) and in Paris (1926–1927), as a pupil of Meillet and Vendryes. His first book, *Principes de grammaire générale*, was rejected as a dissertation because of its synchronic approach. He received his doctorate degree from the university of Copenhagen in 1932 with a volume of studies in Baltic linguistics. From 1934 to 1937, he was professor in Aarhus, then in Copenhagen (chair for comparative linguistics). Among the founders of the Linguistic circle of Copenhagen, he also founded the journal *Acta Linguistica* together with Viggo Brøndal. He died in Copenhagen on May 30, 1965.

References

Hjelmslev, Louis. 1953. Prolegomena to a theory of language. Baltimore: Waverly Press (first published 1943).
———. 1959. Essais Linguistiques. Travaux du Cercle Linguistique de Copenhague, Vol. 12.
———. 1963. Sproget. En introduktion. Charlottenlund: The Nature Method Center.
———. 1973. Essais Linguistiques II. Travaux du Cercle Linguistique de Copenhague, Vol. 14.
Lepschy, Giulio. 1996. Twentieth century linguistics, Vol. 5, History of linguistics, ed. by G. Lepschy. London: Longman.
Malmberg, Bertil. 1983. Analyse du langage au XXe siècle. Théories et méthodes. Paris: P.U.F.
Siertsema, B. 1955. A study of glossematics. The Hague: Mouton.
Spang-Hanssen, H. 1961. Glossematics. Trends in European and American linguistics 1930–1960, ed. by C. Mohrmann, A. Sommerfelt, and J. Whatmough. Utrecht-Antwerp: Spectrum.
Togeby, K. 1965. Structure immanente de la langue française, 2nd edition. Paris: Larousse.
Togeby, K. (ed.) 1967. La glossématique. L'héritage de Hjelmslev au Danemark. Langages 6.
Uldall, H. J. 1957. Outline of glossematics. General theory. Travaux du Cercle Linguistique de Copenhague 10.

SILVIA LURAGHI

See also **Saussure, Ferdinand de; Structuralism**

Hmong and Hmong-Mien (Miao-Yao) languages

Hmong-Mien is a language family and broad cultural grouping of people primarily of China, some groups having emigrated over the last two centuries to Laos, Thailand, Vietnam, and Burma, and in the last 30–40 years as refugees to France, French Guyana, Australia, and the United States of America. There are widely differing uses of terms designating various languages, dialects and cultural groups within this broad grouping—from conflicting uses of everyday terms such as 'Blue/Green Mong', 'Flowery Miao', etc., through to the inclusion of different sets of people and/or languages under the term 'Hmong-Mien' itself, and an alternation between this term and 'Miao-Yao'. The latter term is generally preferred in China, while many Hmong outside of China reject the term 'Miao' as offensive (due to association variously with wild plants as a metaphor for barbarians, or, by an apparently orthographic route, cats). Two consequences of this are that with regard to the people, 'Hmong' is sometimes used to refer specifically to the Hmong outside of China, while in western linguistics literature, 'Hmong-Mien' is becoming the more generally accepted term for the language family as a whole.

Figure 1 shows the genealogical relationships generally accepted for the language family, together with some of the more common alternative language or dialect names. These languages or dialects are for the most part mutually unintelligible, and each subdivides further into a continuum of distinct varieties. Hmong-Mien is typically classified as a Sino-Tibetan language group by Chinese linguists, while western scholarship, debating possible affiliations rather with Austro-Asiatic, tend to leave the question as unresolved to date.

There are around 321,000 speakers of Bunu languages, predominantly Bu-Nao. Note that Bunu speakers are culturally Mien (Yao), and at least another 180,000 Bunu people speak languages other than Bunu.
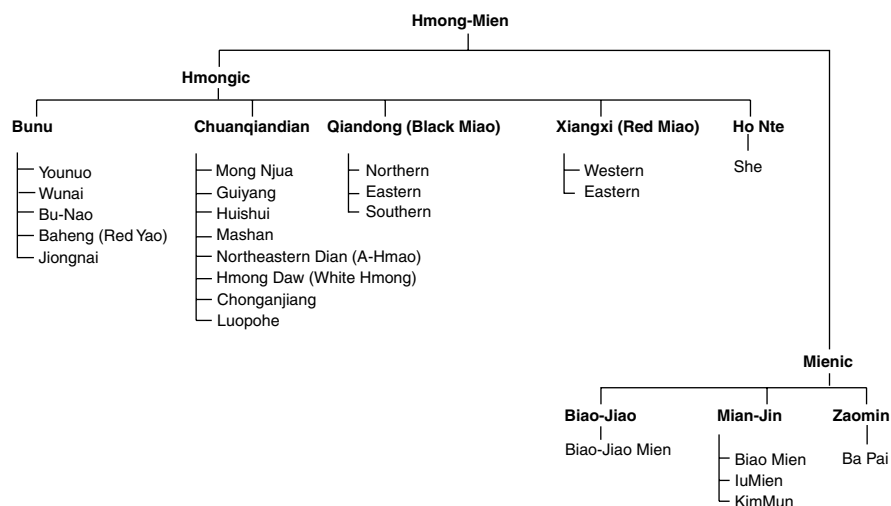
Figure 1. Hmong-Mien family tree.

Qiandong speakers number around 1.4 million, mostly the Northern dialect, while the vast majority of Xiangxi's 770,000 speakers speak the Western dialect, and only a very small number of cultural She are also speakers of She. Chuanqiandian Hmong, something of a repository category for languages that do not fit easily into other subfamilies, boasts 3.6 million speakers, who constitute the bulk of Hmong living outside China. Of these, Hmong Daw and Mong Njua (with around 1.4 million speakers between them) are the primary languages of the ex-refugee diaspora in the west (well over 100,000, mostly in the United States of America). Mong Njua is the most common Chuanqiandian Hmong language in China, with around 1 million speakers, followed by Northeastern Dian with 200,000. Of the 1.7 million speakers of Mienic languages, 1.3 million speak Iu Mien.

Typological Characteristics

Hmong-Mien languages share a range of characteristics, many of which are also evident in other languages of the Southeast Asian region. The discussion below focuses primarily on Chuanqiandian Hmong languages, in particular Hmong Daw, to exemplify some of the language structures typical of this family. (Examples are from Hmong Daw except where otherwise specified. For notes on the RPA writing system used in Hmong Daw and Mong Njua examples, see the section on Writing below. Examples in slash brackets are notated in the International Phonetic Alphabet.)

Syllable Structure of C(onsonant) V(owel)
Hmong-Mien languages typically have a large assembly of initial consonants, incorporating features such as prenasalization, preglottalization, and aspiration. For instance, Baheng includes a consonant series /m mʰ mʲ mʰʲ/, while Xiangxi Hmong consonants include /t tʰ ⁿt ⁿtʰ/.

Although **CV** is the basic syllable structure, some languages also allow final nasals. In Northern Qiandong Hmong, for instance, the vowels /a/ and /o/ may be followed by a nasal /aŋ, oŋ/, while in Mong Njua, nasals always follow a long (and nasalized) vowel, as in the language name 'Mong' /mɔ̃ŋ/.

Lexical Tone
Lexical tone refers to the relative pitch or pitch contour of a given word, when it is part of how that word and its meaning are distinguished. A change in pitch signals a different word, in the same way as a change in a consonant or vowel would change the word. In Baheng, for instance, the word /ko33/ (pronounced with a mid level tone) means 'nine', while /ko22/ (mid-low level tone) means 'road'.

Hmong-Mien languages have variously developed seven, eight, or possibly as many as 13 lexical tones, all of which can be traced back to four original tones in the reconstructed ancestor language (known as Proto-Hmong-Mien). Most languages also display *tone sandhi*—that is, a change in the tone of a given word under specific circumstances, typically in response to the tones of surrounding words. Qiandong Hmong is a notable exception.

Tendency Toward a Monosyllabic Word Structure
This characteristic is necessarily associated with certain principles of grammatical structure and word formation.

Firstly, the words of Hmong-Mien languages do not accrue prefixes or suffixes to indicate meanings such as plurality, possession, case, or the comparative. Rather, where such meanings are overtly present, they are expressed by additional words in the sentence, or by word order. These two strategies, respectively, are demonstrated for the possessive in Example 1:

Example 1

Daim	*dawb*	*yog*	*kuv*	*li*
CLassifier	white	be	I	POSSessive

'The white one is mine'

Kuv tsis		*xav kom*		*kuv tus ntxhais chim*
I	NEGative	want	CAUSative	I CL daughter upset

'I don't want to upset my daughter'

Secondly, the primary strategy for word formation is compounding. This is evident both in long-established words in the languages, and in new words developed in the context of immigration to the west:

Example 2

qab	*qab*	*tsev*	*kawm*	*ntawv*
sweet	sweet	house	study	letters
'delicious'		'school'		

hnub	*chiv*	*nub*	*hli* (Mong Njua)
day	begin	day	moon
'Monday'		'Monday'	

(two of several alternatives current in the United States of America)

Noun Classifiers

This is a set of words that organize nouns into classes, such as 'round things', 'animate things', etc. Their presence or absence with a noun in a given sentence depends on the function they serve—or under certain conditions, they may stand in place of the noun. Classifier functions include clarifying the meaning of the noun, in a way analogous to the use of radicals in Chinese characters, indicating definite rather than nonspecific reference, or they may be associated with enumeration of the noun:

Example 3

Specifying meaning:
daim	*ntawv*	*tsab*	*ntawv*
CL (flat)	paper	CL (written)	paper
'piece of paper'		'letter'	

With numeral:
Thov	*ib*	*rab*	*diav*
please	one	CL (tools)	spoon
'Please could I have a spoon?'			

In place of the noun:
Puas	*muaj*	*tus*	*pab koj ua?*
QUestion	have	CL (animate)	help you do
'Is there someone to help you?'			

Two- and Four-syllable Expressives

These poetic phrases, including but by no means restricted to onomatopoeia, display a distinctive, internally repetitive form, and are scattered throughout everyday conversation, as well as in more literary styles:

Example 4

plib pleb	(of wood crackling)
dig dug	(of the ponderous bubbling of thick liquid)

kev	***noj***	***kev***	***haus***	*nyuaj*	*heev*
way	eat	way	drink	difficult	EMPHatic

'Life is very difficult'

Serial Verb Construction

Hmong-Mien languages feature strings of verbs or verb phrases, which cumulatively express a complex verbal meaning—representing, for instance, an analytic breakdown of minutely differentiated components of an action, or different simultaneous roles of participants in the event.

In the following example, the three verbs *npuav*, *nqa*, and *dua* are combined to describe the complex action of carrying something somewhere. Also note that the action of picking up the cubs must first be specified by yet another verb:

Example 5

Ob	*tus*	*dais*	*muab*	*cov*	*menyuam*
two	CL	bear	take	CL	child
npuav	*nqa*	*dua*	*nram*	*kwj-hav*	
carry	carry	pass	down	valley	
by mouth					

'The two bears picked up their cubs and carried them down into the valley'

Reliance on Context to Determine Functions Such as Tense and Plurality

A range of grammatical and pragmatic functions are not obligatorily specified in languages of this family. In the first of the following examples, the tense is implied by the aspect marker *tau* (attainment), while in the second, it can be understood only in context:

Example 6

Yuav	*tau*	*ib*	*lub*	*hnab-looj-tes*
buy	ASPect	one	CL	gloves

'I bought a pair of gloves'

Kuv	*yuav*	*lub*	*hnab-looj-tes*
I	buy	CL	gloves

'I am buying/bought/will buy some gloves'

Also note also the omission of the pronoun in the first example, and the assumed plurality of *hnab-looj-tes*.

Concluding Remarks

The above sample characteristics of Hmong-Mien languages are not a random combination. Rather, certain characteristics, such as underspecification, classifiers, and serial verb constructions, co-occur in linguistically logical and historically traceable combinations. This, in part, explains the commonality of these and other characteristics to not only this language family but also others of the region. Other aspects of Hmong-Mien languages are specific to that family—for instance, the ordering **Possessor–Possessed** illustrated in Example 1 above.

Writing

Writing and orthography are a key issue for many Hmong-Mien groups, and a good number of writing systems have been proposed and to varying extents have been implemented—either from within the communities, or by missionary linguists working with the communities, or by government organizations of the various countries of residence. The question is not only one of practical literacy in a heritage language, but is also, at least for the Hmong, embedded in cultural narratives of national identity and mythic histories, or *dab neeg*. In these *dab neeg* are encoded a Hmong history of oppression, migration, and ancient literacy, which is in turn linked to strong intragroup cohesiveness and long-cherished goals of self-determination.

Writing systems have been developed on a base of Chinese (e.g. Xiangxi and Qiandong Hmong), Thai, and Lao scripts, as well as the Roman—Vietnamese base, Pinyin base (e.g. Mien languages), and Roman Popular Alphabet (RPA—Hmong Daw and Mong Njua). In addition, there are at least four independent systems: Pollard script (Northeastern Dian), Pahawh Hmong (Hmong Daw and Mong Njua), *Ntawv Paj Ntaub* (Hmong Daw and Mong Njua, used only in a small area of California), and the *Ntawv Puaj Txwm* (an apparently ancient script).

The most widespread writing system now in use is the RPA. This is based on a featural representation of the sounds of Hmong, and utilizes consonant symbols syllable-finally to indicate tone. The word 'Hmong Daw'/m̥ɔ̃ŋ⁵⁵ daɨ⁵⁵/ is written <Hmoob Dawb>:

Example 7

hm	oo	b	d
preaspirated or voiceless nasal	long vowel (nasalised)	high level tone	alveolar stop
aw	b		
diphthong	high level tone		

(Note that the final nasal of 'Hmong' is not directly notated, being a predictable closure of the long vowel.)

Other tones are written as:

Example 8

m	low glottalized	**j**	high falling
v	mid rising	**d**	low rising, lengthened
s	mid-low falling	**g**	low breathy
(zero)	mid level		

The Pahawh Hmong, although restricted in use for the most part to small communities located throughout Australia, the United States of America and in northern Thailand, is recognized by many as a significant symbol of Hmong autonomy and, particularly for its proponents, as a realization of the millenarianist prophetic tradition threaded throughout the *dab neeg*.

The Pahawh is based on the **CV** syllable structure of Hmong, so that each syllable requires only one consonant and one vowel grapheme. A limited set of diacritics combines with each grapheme type, to (i) form a vowel-tone composite graph, or (ii) specify one of three consonants possible for a given grapheme—reducing the number of graphemes that would otherwise be needed for 59 consonants (in Hmong Daw). Example 9 is in Pahawh Version Two, as current in major Pahawh teaching organizations. (Version Three, also found in older and linguistics texts, is not fully readable on the basis of this Version.) Note that, although the writing is from left to right, the order within a word is from right to left (Vowel+tone–Consonant):

Example 9

Pahawh	ả̌	ṋ̄	ᕜ̋
RPA	*Phaj*	*hauj*	*Hmoob*
			'Pahawh Hmong'

This example demonstrates the use of zero diacritic < , >, representation of the same tone by different diacritics depending on the base vowel grapheme < ả , >, and similarly use of the same diacritic for different consonants, depending on the base consonant grapheme < , >. Many features of Pahawh writing, including word-internal ordering and the shapes of the graphemes themselves, hold symbolic significance in the tradition of its emergence.

The Hmong who have resettled in the west after the refugee period have taken advantage of the situation by embarking on a new period of accelerated language development. This includes not only orthographic development but also ongoing creation of a wealth of new words, in order to ensure the useability of the heritage language(s) in all the domains now encountered by the rising generation in the cultures of immigration. A third area of development has been a surge of publications in Hmong (Hmong Daw and Mong Njua)—language teaching and reference

materials including dictionaries and literacy primers, books on culture and history both traditional and emerging, community magazines, novels, and a rapid proliferation of websites and related communications media.

References

Bisang, Walter. 1993. Classifiers, quantifiers and class nouns in Hmong. Studies in Language 17(1). 1–51.

Bisang, Walter. 1996. Areal typology and grammaticalization: processes of grammaticalization based on nouns and verbs in east and mainland South East Asian languages. Studies in Language 20(3). 519–97.

Court, Christopher. 1986. Fundamentals of Iu Mien (Yao) grammar. Ph.D. Thesis. Berkeley: University of California.

Edmondson, Jerold. 1992. Pa-hng development and diversity. Papers from the first annual meeting of the Southeast Asian Linguistics Society 1991, ed. by Martha Ratliff and Eric Schiller. Tempe, AZ: Program for Southeast Asian Studies, Arizona State University.

Eira, Christina. 2002. Language maintenance at the micro level: Hmong ex-refugee communities. Language endangerment and language maintenance, ed. by David Bradley and Maya Bradley. London and New York: RoutledgeCurzon.

Enwall, Joakim. 1994. A myth become reality: history and development of the Miao written language, Vol. 1. Stockholm: Institute of Oriental languages, Stockholm University.

Grimes, Barbara. (ed.) 2000. Ethnologue, Vol. 1: Languages of the world, 14th edition. Dallas: SIL International.

Mottin, Jean. 1978. Eléments de grammaire Hmong Blanc. Bangkok: Don Bosco Press.

Ratliff, Martha. 1992. Meaningful tone: a study of tonal morphology in compounds, form classes, and expressive phrases in White Hmong. Northern Illinois: Center for Southeast Asian Studies, Northern Illinois University.

Smalley, William, Chia Koua Vang, and Gnia Yee Yang. 1990. Mother of writing: the origin and development of a Hmong messianic script. Chicago: University of Chicago Press.

Tapp, Nicholas. 1989. Sovereignty and rebellion: the White Hmong of northern Thailand. Singapore: Oxford University Press.

Wang, Fushi. 1985. Miaoyu jianzhi (Sketch of the Miao language). Beijing: Renmin Chubanshe.

CHRISTINA EIRA

See also **China**

Hockett, Charles F.

Charles F. Hockett, one of the leading neo-Bloomfieldian linguists in the United States, authored more than 130 publications covering a broad spectrum: general linguistics, phonology, morphology, semantics, sound change, historical linguistics, linguistic universals, the origin of language, animal communication, Chinese, Fijian, and analyses of such American Indian languages as Potawatomi, Ojibwa, and Arapaho. As a student of Edward Sapir and Leonard Bloomfield, Hockett was trained both as a linguist and an anthropologist and never wavered from his conviction that the disciplines are inseparably intertwined. At the height of his remarkable career in the late 1950s, Hockett's behaviorist view of language and structuralist approach to linguistic analysis was eclipsed by transformational grammar, Noam Chomsky's revolutionary linguistic theory that attacked the very foundations of structuralism: behaviorism and the neglect of semantics. Hockett rejected the validity of transformational grammar and continued publishing spirited rebuttals in defense of his own views on the nature of language for two decades. As the literary executor of his mentor, Leonard Bloomfield, he devoted much time and energy publishing posthumously the bulk of Bloomfield's research material on Menomini, an American Indian language, and compiling an anthology of his most important publications.

In the 15 years between Bloomfield's death and Noam Chomsky's transformational grammar, Hockett was one of the leaders of American linguistics known especially for his role in the development of American phonological theory. As a post-Bloomfieldian structuralist, Hockett was convinced that, as he later said, linguistic analysis must be done 'from the bottom up' beginning with the sound system, progressing to the grammatical level of inflection and word formation, and only then to the syntactic level, the rules governing the arrangement of words in the sentence. Each of these levels of language was to be kept strictly separate. This meant that grammatical considerations could not be used to determine the distinctive sounds of a language, the phonemes. European phonologists of the Prague school viewed the phoneme as consisting of a set of features that could partially overlap with those of another phoneme, so that the distinction between two phonemes could be neutralized in certain positions. The phonemes /p/ and /b/ are distinguished by the feature of voicing, /p/ being voiceless, /b/ being voiced. If, as is the case in German, /b/ is

always pronounced as /p/ at the end of a word, the contrast between these two phonemes is neutralized in the word-final position. American structuralists, Hockett foremost among them, considered each phoneme to be unique and rejected partial phonemic overlap or neutralization outright. Together with the strict separation of linguistic levels, the requirement 'once a phoneme, always a phoneme' became a defining characteristic of American structuralism and a mark of its theoretical rigidity. One of Hockett's reasons for rejecting transformational grammar was the 'top-down' approach to linguistic analysis beginning with syntax and semantics rather than phonology. According to Hockett, linguistic analysis from the top-down is inappropriate because the investigator has to isolate words or morphemes before work on syntax and semantics can begin.

At the peak of his career, Hockett was perhaps best known for his textbook *A course in modern linguistics* (1958), which, as the author states in the introduction, was not written to explore frontiers and indulge in flights of fancy, but to present the generally accepted facts and principles of the field. He also included chapters on dialectology, historical linguistics, and man's place in nature. He compared animal communication systems, for example, the dance of honey bees, with human language by means of seven key properties. The property that, according to Hockett, is unique to human language is duality of patterning, a term Hockett invented to account for the fact that an infinite number of meaningful forms are built from a small number of meaningless elements.

Hockett's monograph *The state of the art* (1968) is a spirited attack on the basic principles underlying transformational grammar. As a behaviorist who insisted that only observable utterances could be the subject of linguistic analysis, he questioned Chomsky's distinction between a speaker's linguistic *competence*, the innate knowledge of the rules of his or her language, and *performance*, the speaker's actual utterance, which may be flawed by mispronunciations, inadvertent errors, repetitions, and the like. For Hockett, performance was the subject of linguistic investigation; for Chomsky, it was competence. He also disagreed with the Chomskyan premise that a speaker has access to an infinite number of sentences of infinite length and argued that there are limitations to sentence length that are part of the structure of language itself. When *the State of the Art* was published, it was already clear that transformational grammar had won the day, Hockett's spirited rearguard action notwithstanding.

Hockett's most significant and lasting contributions are, perhaps, his many publications on American Indian and Austronesian languages, and his role as Leonard Bloomfield's literary executor.

Biography

Charles Hockett was born on January 17, 1916 in Columbus, Ohio. He entered Ohio State University at 16 years of age. He received a combined B.A. and M.A. degree in ancient history in 1936. He worked as graduate student in anthropology at Yale with the anthropologist/linguist Edward Sapir and the structuralist linguist Leonard Bloomfield. Hockett received his Ph.D. in anthropology in 1939. He was drafted during Word War II and wrote an elementary Chinese text book for the War Department. Hockett joined the linguistics faculty at Cornell University in 1946 as an assistant professor and was promoted to Professor of Linguistics and Anthropology in 1956. In 1955–1956, he received a fellowship at the Center for Advanced Studies in the Behavioral Sciences at Palo Alto, California. He was president of the Linguistics Society of America, 1964–1965 and became Goldwin Smith Professor of Anthropology and Linguistics in 1970. In 1974, he was elected to the National Academy of Sciences. After retirement in 1982, he served as an adjunct professor at Rice University. He died on November 3, 2000 in Ithaca, NY.

References

Agard, Frederik B., et al. (eds.) 1983. Essays in honor of Charles F. Hockett. Leiden: Brill.

Chevillet, Francois. 1996. The phoneme and American linguistics: an interview with C.F. Hockett. Etitudes Anglaises 49. 180–91.

Fillmore, Charles J. 1969. Review of the state of the art. American Anthropologist 71. 711–3.

Harris, Randy Allen. 1993. The linguistics wars. Oxford: Oxford University Press.

Hocket, Charles F. 1942. A system of descriptive phonology. Language 18. 3–21.

———. 1954. Two models of grammatical description. Word 10. 210–34.

———. 1955. A manual of phonology. Indiana University Publications in Linguistics and Anthropology, Memoir 11.

———. 1958. A course on modern linguistics. New York: Macmillan.

———. 1962. Posthumous publication, ed. by L. Bloomfield. The Menomini language, New Haven: Yale University Press.

———. 1968. The state of the art. The Hague: Mouton.

——— (ed.) 1970. A Leonard Bloomfield anthology. Bloomington: Indiana University Press.

———. 1973. Man's place in nature. New York: McGraw-Hill.

———. 1974. Posthumous publication, ed. by L. Bloomfield. Menomini lexicon. Milwaukee Public Museum Publications in Anthropology and History, no. 3.

———.1977. The view from language. Selected essays 1948–1974. Athens: University of Georgia Press.

———. 1980. Preserving the heritage. First person singular, ed. by Boyd H. Davis and Ramond O'Cain. Studies in the History of Linguistics, Vol. 21. Philadelphia: John Benjamins.

———. 1987. Refurbishing our foundations. Elementary linguistics from an advanced point of view. Philadelphia: John Benjamins.

Joos, Martin. (ed.) 1957. Readings in linguistics I. The development of descriptive linguistics in America 1925–56. Chicago: The University of Chicago Press.

RENATE BORN

See also **Bloomfield, Leonard; Sapir, Edward**

Humboldt, Wilhelm von

The most distinguished linguistic scholar of the nineteenth century, politician, diplomat, and humanist, Karl Wilhelm Freiherr von Humboldt is generally considered to be the founder of general linguistics as a science that aims at the quest for the essence of linguistic phenomenon itself through the analysis and comparison of different languages. Together with his younger brother, Alexander von Humboldt, a natural scientist, he was given privileged education in classical studies by excellent private tutors that not only resulted in his profound and life-long interest in languages, philosophy, and humanities in general, but also motivated his later work on reforms of the Prussian educational system and the foundation of the University of Berlin. In addition to Greek and Latin, he studied Sanskrit and other Indo-European and non-Indo-European languages such as the ancient Indonesian language Kawi and Basque, while his brother Alexander, a pioneering explorer of the Americas, often sent back to him data collected from Amerindian languages with which he came into contact.

As a contemporary of German philosophers Immanuel Kant, Friedrich Hegel, and Johann Gottfried von Herder, and the Romantic writers Wolfgang Goethe and Friedrich von Schiller, Humboldt synthesized the main ideas of philosophical idealism and the Romantic emphasis on free, individual creativity and subjective meaning into his philosophy of language. He outlined his full theory in a lengthy introduction to his life work on the Jawan language Kawi, *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschlichengeschlechts* (1836; *On language: the diversity of human language—structure and its influence on the mental development of mankind*). Based on the Kantian epistemological stance of cognition as activity, through which mental categories were imposed upon sensible experience, Humboldt held that language was not only a means of communication but also an a priori framework of cognition through which one formulated one's thoughts. According to Humboldt's famous dictum, language in its very essence is not a product (*ergon*) but a process and activity (*energeia*) inherent to human nature, an organ that forms the thought. Its true definition can therefore only be a genetic one, with the 'form of language' as largely programmed by the internal disposition of a genetic program. It is the 'inner form', the pattern, or structure, of grammar and meaning that is imposed on the 'outer form', the sounds from which different languages are differentiated one from another. This repeated work of the mind enables the articulated sounds to express thought, which makes an inseparable unity between language and thought and language and sounds. In accord with his theory, Humboldt holds that language acquisition does not occur through mechanical learning but is a matter of development of cognitive capacities under appropriate external conditions, just as all main human capacities develop at certain definite growth stages of life.

Postulating that reason is inseparable from language, thought is inseparable from speech, Humboldt infers, in accordance with his dynamic view of language as activity, that languages are not really means for representing already known truths but are rather instruments for discovering previously unrecognized ones. However, from his metaphysical concept of the 'inner form', he further concludes that thought does not depend only on the universal human capacity of speech but also on the different realizations of this 'inner form' in different languages. This blend of relativistic concept of the language role in intellectual processes and the idealistic, universalistic interpretation of the essence of the linguistic phenomenon by means of conceptual categories, represents the theoretical basis of Humboldt's deterministic doctrine on language as a worldview (*Weltanschauung* theory). Emphasizing the difficulties of full understanding in the process of communication and translation from one language into another, Humboldt indicates that

'each language sets certain limits to the spirit of those who speak it; it assumes a direction and, by doing so, excludes many others'. Moreover, language acts not only as an intermediary between man and the world but also between the individuals in a community and the whole humankind, so that in social terms, differences in individual linguistic worldviews lead to differences in group worldviews. Following the views of Herder, Humboldt maintains that each language and each culture reflect the world in a particular way and that the mental quality of each nation and its culture determines the language of its people, which in its turn determines the way they think and experience reality.

Analyzing a large number of world languages by a comparative method, Humboldt also made original contributions to language typology. His studies on the evolution of language and classification of languages, however, were closely related to his view on interdependency between language and the mental spirit of a nation or culture. He based his work on Friedrich Schlegel's genealogical scheme, which distinguished between the most developed 'inflectional' languages, i.e. the Indo-European languages, and those that have no inflection and are therefore called 'isolating' (e.g. Chinese) or the intermediary type of the so-called 'agglutinating' languages (as Amerindian), which put strings of forms together, but do not allow for a modification of the root. Somewhat controversially, Humboldt believed that these differences in language types reflected the developmental stages of human intellect, claiming that the highest achievement of the human mind was that of the speakers of highly inflectional Ancient Greek. Although he did not relate these types to historical progress or developmental stages of culture, he has been sometimes criticized for implying the idea of superiority of one people over another, based on the differences of language structure.

Humboldt's complex and rich ideas on dynamic language concept, 'inner form', and linguistic worldview have continuously influenced the subsequent course of linguistics in both Europe and the United States. In Europe, his legacy is evident not only in the work of his immediate successors, such as Franz Bopp, Heymann Steinthal, and the forerunner of psycholinguistics Wilhelm Wundt, and of the later group of the so-called Neo-Humboldtians, including Leo Weisgerber and Jost Trier, who advocated the majority of Humboldt's theoretical ideas, but also in some basic concepts of Ferdinand de Saussure's structural linguistics. In American linguistics, the essentials of his linguistic relativity and *Weltanschauung* theory found their clear expression in the works of the founders of linguistic anthropology Franz Boas, Edward Sapir, and Benjamin L. Whorf, while his distinction of inner and outer linguistic form represents the basic notions of generative grammar elaborated by one of the most influential modern linguists, Noam Chomsky.

Biography

Wilhelm von Humboldt was born in Potsdam, Prussia on June 22, 1767. He studied law at the University of Brandenburg in Frankfurt (Oder) in 1787, moved to Göttingen University where he also began studies in classical philology in 1788. He joined Prussian civil service as a law clerk to the Supreme Court of Berlin in 1789 and resigned from civil service in 1790. He resided on family estates in Thuringia and Jena and there he continued his studies with private tutors, devoted his interests to literature criticism and esthetics, and associated with Goethe, Schiller, and the Schlegel brothers between 1790 and 1797. He moved with his family to Paris to observe the turbulent social development in France in 1797–1799, and then to Spain where he investigated the Basque language. He rejoined the Prussian civil service and became an envoy to the Vatican in Rome in 1802–1808. He was head of the Department of Education and Arts at the Home Office in Berlin, and reformed Prussian educational system by elevating standards for teachers' training and founded the Friedrich Wilhelm University (modern Humboldt University) in Berlin, 1809. In 1809–1812, he retired to private life. He re-entered diplomatic service as the Prussian ambassador in Vienna and later London in 1812–1819. He resigned from civil service in 1819 and continued his private philological studies in the quiet atmosphere of the family manor in Tegel, investigating the Basque, Sanskrit, American, and Oriental languages and writing his most important linguistic works on the philosophy of language. He died in Tegel, Prussia on April 8, 1835.

References

Aarsleff, Hans W. 1988. Introduction. On language—the diversity of human language—structure and its influence on the mental development of mankind. New York: Cambridge University Press.

Basilius, Harold. 1968. Neo-Humboldtian ethnolinguistics. Readings in the sociology of language, ed. by Joshua Fishman. The Hague: Mouton.

Brown, Roger L. 1967. Wilhelm von Humboldt's conception of linguistic relativity. The Hague and Paris: Mouton.

Cowan, Marianne (ed.) 1963. Humanist without portfolio: an anthology of the writings of Wilhelm von Humboldt. Detroit:Wayne State University Press.

Humboldt, Wilhelm von. 1821. Über den Einfluss des verschiedenen Charakters der Sprachen auf Literatur und Geistesbildung; Linguistic variability & intellectual development, transl. by George C. Buck and Frithjof A. Raven, 1971.

———. 1836. Über die Verschiedenheit des menschlichen Sprachbaues undihren Einfluss auf die geistige Entwicklung des Menschengeschlechts. Berlin: Dümmler; as Linguistic

variability and intellectual development, transl. by George C. Buck and Frithjof A. Raven, Coral Gables, FL: University of Miami Press, 1971; as On language: the diversity of human language and its influence on the mental development of mankind, transl. by Peter Heath, Cambridge: Cambridge University Press, 1988; as On language: on the diversity of human language construction and its influence on the mental development of the human species (Cambridge texts in the history of philosophy), ed. by Michael Losonsky, transl. by Peter Heath, Cambridge: Cambridge University Press, 1999.

———. 1903–1936. Gesammelte Schriften (Collective works), ed. by von Albert Leitzmann, Bruno Gebhart and Wilhelm Richter, Königlichen Preussischen Akademie der Wissenschaften, 17 vols., Berlin: Behr; Berlin: De Gruyter, 1967–1968.

———. 1963. Humanist without portfolio: an anthology of the writings of Wilhelm von Humboldt, ed. and transl. by Marianne Cowan. Detroit: Wayne State University Press.

———. 1997. Wilhelm von Humboldt: essays on language, ed. by Theo Harden and Dan Farelly, transl. by John Wieczorek and Ian Roe. Frankfurt and New York: Peter Lang.

Manchester, Martin L. 1985. The philosophical foundations of Humboldt's linguistic doctrines. Amsterdam and Philadelphia: Benjamins.

Sweet, Paul R. 1978. Wilhelm von Humboldt: a biography, 2 vols. Columbus, OH: Ohio State University Press.

ANITA SUJOLDZIC

Hungarian and Ugric Languages

Hungarian together with Mansi (Vogul) and Hanty (Ostyak) belongs to the Ugric group of the Finno-Ugric language family. In its turn, Ugric group is subdivided into Hungarian subgroup and Ob'-Ugrian subgroup, containing Mansi and Hanty. Hungarian is the official language of the Hungarian Republic (Magyar Nepkoztarsasag). The Hungarian language (magyar nelv) is spoken by Hungarians or Magyar, as they call themselves. In other languages, they have different names, e.g. German 'Ungarisch', Ukrainian 'Ugor'ska', Russian 'Vengerskiy', etc. A certain number of Hungarians live in the southwest of the Ukraine (Uzhgorod area), in Romania, Slovakia, Canada, and the United States of America. The total number of speakers may be estimated as 12–14 million.

Hungarians are said to have arrived to their present positions around the lake of Balaton from their original homeland in the South of Western Siberia and the Urals. Actually, Mansi and Hanty still live in the Northern Urals and Western Siberia. The Old Hungarian tribes moved to the West from the Urals and lived for about a thousand years on the Volga and Oka river basins where they had close contacts with the Permian peoples (Komi) as well as the peoples of the Volga region: Mari (Cheremis), Mordva, and the ancient Bulgars, the tribes of Turkic origin, who partly moved to the lands of modern Bulgaria, partly merged with the Tatars, Chuvash, and partly joined the ancient Hungarians. After living for some time on the Don river, the ancient Hungarian tribes moved again westward to conquer the territory of the Danube and Balaton basins from the Slavonic tribes of Slovaks in the ninth century AD. The ancient Hungarian king Arpad completed their 'settlement' (honfoglalas) by AD 896. In the fifteenth century, the flourishing period of national formation was interrupted by the Turkish invasion and then the Austrian rule. Only since 1918, after the collapse of the Austrian–Hungarian Empire, did Hungarian begin to develop without strong foreign influences. These numerous language contacts influenced modern Hungarian enormously, leading to the gap between Hungarian and Ob'-Ugrian languages (Mansi and Hanty). Now, communication between a Hungarian, Mansi, and Hanty is impossible. The oldest Hungarian literary relics are the 'Funeral oration' (Halotti Beszed, c. AD 1200) and 'Lament of Mary' (o - Maria Siralom), which is dated about a hundred years later. The greatest Hungarian poets are Sandor Petofi, Janos Arany, and Endre Ady. (In Hungarian it should be Ady Endre, because the family name comes first followed by the first name of a person.) Hungarian uses Roman alphabet without some letters (q, x, y) but with a system of diacritics for vowel length and quality. Long vowels are shown by ['] over a letter, e.g. a. Two dots over [o] and [u] like o and u shows that they are pronounced mildly, that is, like 'Umlaut' in German. The sign ['] shows long vowels with umlaut, e.g. o. Many consonant sounds are given by combinations of letters, e.g. cs [t?], dz [dz], dzs [dz], gy [d'], ly [l'], ny [n'], sz [s], ty [t'], zs [z]. Hungarian also has long consonants. The length of the consonant is denoted by doubling the letter, e.g. pp, bb, dd, mm, ll, etc. It is phonemic, i.e. they change the meaning of the word: hal 'fish' —hall 'hears'; var 'wound' —varr '(he) sews'. Long consonants built by letter combinations require doubling

only the first consonant letter, e.g. ccs, ddzs, ggy, ssz, nny, zzs. One can see that the letter 'y' is not pronounced. It shows that the preceding letter is milder, e.g. gy [d'], ny [n'], ly [l'], ty [t'].

The phonology of Hungarian is rather simple. It has no diphthongs, but it has seven long vowels denoted by ['] over the vowel. In the phonetic transcription, we use [:], e.g. a [a:]. There are 14 vowels all in all: [i, i:, u, u:, u, u:, e, e:, o, o:, o, o:, a, a:], among them eight are labialized [o, o:, u, u:, o, o:, u, u:]. According to the vertical position of the tongue, three types of vowels are defined: high vowels: [i, i:, u, u:, u, u:], middle [e:, o, o:, o, o:], and low [e, a, a:]. According to the horizontal movement of the tongue, we define two types of vowels: front [i, i:, u, u:, e, e:, o, o:] and back [a, a:, o, o:, u, u:]. The length of the vowel is phonemic, i.e. it changes the meaning of a word: lap 'page', lap 'marsh, bog', etc. Like in Turkish, Tatar, and other Turkic languages, there is vowel harmony in Hungarian. It means that in a Hungarian word, all vowels share a certain articulation pattern. They are produced with either the front or the back of the tongue, e.g. the locative suffix '-ban' is not changed when there is a back vowel in the word root:_varos 'town' - varosban - 'in town'; a haz 'the house' - a haz+ban 'in the house'; haz+hoz 'to the house'. However, the suffixes change if the root vowel is in front, e.g. a viz 'the water' - a viz+ben 'in the water'; kert+hez 'to the garden'. The stress is always on the first syllable. If the word is long, then it is also on the third syllable.

The hungarian consonantal system is rich. There are 50 consonants in Hungarian all in all. It is 3.6 times greater than the number of vowels. We can define them by the work of the active organ of speech (i.e. the place of articulation), the manner of articulation (i.e. the type of obstruction), and the work of the vocal cords. According to the work of the active organ of speech, we define the following consonants: labio-labial [p, p:, b, b:, m, m:], labio-dental [f, f:, v, v:] (for the lack of place, we shall not show the long vowels further); front (alveolar and dental) [t, d, s, z, ?, ts, dz t?, z, dz, n, l, r]; palatal (mediolingual) [t', d', j]; back (velar) [k, g, h], pharyngeal [h]. According to the manner of articulation, the following consonants are defined: sonorant [m, n, n', l, r, j]; occlusive [p, b, d, d', t, t', k, g]; affricate [ts, dz, t?, dz]; and fricative [v, s, z, ?, z, h]. According to the work of the vocal cords, we define voiced [b, v, d, d', z, dz, z, dz, g] and voiceless [p, f, t, t', s, ?, ts, t?, k]. As we can see further, the phonetical system of Hungarian is rather different from those of Mansi and Hanty.

There is no grammatical gender in Hungarian. Its morphology is of an agglutinative type, i.e. there are no prefixes before the root; the root does not change; suffixes are added ('glued') to the root, and then to

each other, e.g. haz+am 'my house'; 'haz+ad 'your house'. The plural marker '-k' is added to the consonantal stem according to the rules of vowel harmony a haz+ak 'the houses'; a kep+ek 'the pictures'. Linear addition of suffixes is characteristic of agglutinative languages (e.g. Finno-Ugric and Turkic languages). The Hungarian case system is rich. Nouns, pronouns, adjectives, and numerals are declined according to the same rules. There are 23 cases. Unlike many other languages of the world, in Hungarian, numerals are followed by a noun in the singular, kilenc haz 'nine houses'. Verbs can be transitive and intransitive in the indicative, conditional, or subjunctive [imperative moods. Definite and indefinite conjugation.

Hungarian has only 600–700 words of Finno-Ugric or Uralic origin. All the rest of the Hungarian words are borrowings. Old Hungarian borrowed a small number of words from Iranian (through Old Ossetian, e.g. tej 'milk', tehen 'cow', hid 'bridge', var 'fortress'). The first larger layer of borrowings (about 300 words) came in the fifth to ninth centuries from the Turkic languages (especially Chuvash), which belong to cattle breeding, husbandry, crafts, house building, cloth making, public and spiritual life. The second Turkic word layer entered Hungarian from the Pecheneg and Polovets languages in the tenth to thirteenth centuries. Some part of the Pecheneg and Polovets joined the Old Hungarian at that time. A great number of words (more than 500) were borrowed from the Slavonic languages, especially Slovak and Russian. They pertain to all spheres of everyday life. German also influenced Hungarian greatly giving it about 400 words concerning crafts and sciences. Religious, cultural, and school terms came from Latin. Dozens of words were borrowed from Italian, Romanian, or Spanish. So one can see that the major part of Hungarian is the words of non-Finno-Ugric origin.

The Mansi Language

Mansi belongs to the Ob'-Uric subgroup of the Ugric group of the Finno-Ugric family. Anthropologically, they are said to belong to the Mongoloid type. Some scholars believe them to have migrated to the North of Siberia from the Sayan mountains of the South of Siberia. They were called formerly Vogul, which is not correct because the people call themselves 'Man's'i'. In 1989, there were 8,474 native Mansi, but only 37.1% of them considered Mansi as their native language. May be, half of those 37.1% can actually speak Mansi. They live on the territory of the Hanty-Mansi autonomous national district in Russia, which is situated in the basins of the Ob', Sos'va, and Konda rivers and in the Northern Urals. There were four dialects in Mansi, whose speakers did not understand

each other. These dialects were so different that they could be called different languages. Actually, they were more different than German and English. Now, only some of the Northern Mansi can speak their native language, the rest of the Mansi dropped it and began to talk Russian. Two dialects of Mansi, Northern and Konda, are analyzed here to show how different they are. Russian Orthodox missionaries tried to christen Mansi in the second half of the nineteenth century; therefore, they translated the New Testament into Mansi. It was the first book in Mansi. Literary Mansi languages, based on the Northern dialect, was introduced into the mass of illiterate Mansi in the 1930s only. Then, the script was Latin, which was changed into Cyrillic in 1937. Modern Mansi borrowed many words of everyday life from the language of Siberian Tatars, who ruled Mansi before the Russians came to Siberia in the sixteenth to seventeenth centuries. Mansi has many Russian borrowings in the field of culture, sciences, arts, and education. Mansi is used in the first four forms of schooling and also on the radio and in one newspaper. The phonology of Mansi is much simpler than that of Hungarian. It has six vowels: [i, e, I, a, o, u]. Some scholars consider Mansi to have long vowel counterparts. Vowel length is not depicted in the script. This problem has not been studied by the methods of experimental phonetics therefore, one cannot state for sure that the vowel length is phonemic. It can distinguish only some words with [a] and [o]. Anyway, it needs more investigation. According to the vertical movements of the tongue three positions, thus, three vowel types are defined: high [i, u], middle [e, I, o], and low [a]. According to the horizontal movement of the tongue three positions, thus, three types of vowels are defined: front [i, e], central [I], and back [a, o, u]. Unlike Hungarian, there is no vowel harmony. The stress falls on the first, third, and fifth syllables. The stressed vowels sound longer than the unstressed ones. The number of consonantal phonemes in Mansi is not so great as in Hungarian. There are no long consonants. By the work of the active organ of speech (place of articulation), the following consonants can be defined: labio-labial [p, w, m], (no labio-dental whatsoever!); front [t, s, n, l, r]; palatal [t', s', n', l', j]; and velar [k, x, ?, ?]. By the manner of articulation, the following Mansi consonants are classified into four types: sonorant [m, w, n, n', l, r, j, ?], occlusive [p, t, t', k], (no affricates), and fricative [s, s', x, ?]. By the work of vocal cords, nonsonorant Mansi consonants are divided into two groups: voiced and voiceless. There is only one voiced [?], all the rest are voiceless [p, t, t', s, s', k, x]. In this respect, as in many others, Mansi is quite different from Hungarian. Communication between a Mansi and a Hungarian is impossible. Communication

between Northern and Konda Mansi is also impossible. Actually, Konda Mansi is quite a different language. This can be seen even from its phonemic system. Konda Mansi has 17 (short and long) vowels and 18 consonants. Nowadays, it is an extinct language. Thus, one cannot state for sure if the long vowels are phonematic. There exists a strong doubt. It may be stress that makes Konda vowels long.

According to the work of the active organ of speech, the following four groups of consonants are defined: labio-labial [p, w, m], (there are no labio-dental consonants); front [t, s, ?, l, ?, r]; palatal [t', s', n', l', j]; and velar [k, x, ?, ?]. According to the manner of articulation, the following four groups are defined: sonorant [m, n, n', l, l', ?, ?]; occlusive [p, t, t', k]; (there are no affricates); and fricative [w, s, s', ?, x, ?]. There are only two voiced nonsonorant consonants [w, ?]. One can see that the phonemic nomenclature of Konda Mansi is more similar to Northern Mansi than to Hungarian. Mansi is an agglutinative language. The suffixes are added to the root: hap 'boat'; hap+um 'my boat'; hap+ag+um+t 'in my (two) boats; nepak+ag+men 'our (two) two books'; kat+ag+um+til 'by my (two) hands'; war+s+an+en -' they (two) did these'; and jal+as+al+is 'he traveled many times', i.e. jal- 'to go', '-as' the suffix showing contiguity of action, '-al' the suffix showing the interruption of action; '-is' the suffix showing the past tense, and the lack of personal ending shows the third-person singular. Mansi nouns have no gender. The noun has number: singular, dual, and plural. The Mansi noun has seven cases, Hanty, 10. It takes possessive suffixes and is definite or indefinite. There are a small number of adjectives, but mostly nouns can serve as adjectives if they are in preposition to the other nouns, e.g. ker unsah 'iron bridge'; being in preposition to the noun, it does not agree with the noun in number or case, being a part of the predicate, it agrees with the subject in number.

There are several types of Mansi pronouns: personal, demonstrative, interrogative, negative, and indefinite. Most of them are declined. Mansi, like Hungarian, has definite and indefinite conjugations and five moods: indicative, imperative, subjunctive, conditional, and inferential. Word order is usually S(ubject)-O(bject)-V(erb) or OSV.

Hanty

Hanty together with Mansi form Ob'-Ugric subgroup of the Ugric group (plus Hungarian) of the Finno-Ugric language family. They appear more Mongoloid than Mansi. They call themselves Hanty, although in the West they are still wrongly called Ostjaks. They live in the basins of the Ob' River and Vas'ugan River.

The majority of Hanty live in the Hanty-Mansi national District (11,900 people), although some (7,300 people) live in the Jamal-Nenets national autonomous District or in Tomsk region of Russia. The Hanty language is known for its dialectal diversity; nearly every village speaks differently. Only 67.8% of Hanty consider it their native language. In the 1930s, an attempt was made to set up a literary Hanty language, but the success was even smaller than with Mansi. The Northern (Kazym) dialect was chosen to be the basis of the literary Hanty language. The first desk books were published in that dialect. However, only the Kazym Hanty could use them, while the native speakers of the other dialects could not understand them. This is why, in the 1940s they also began publishing books in the Obdor (Salehard), Ber'ozovo, and Middle Ob' (Sherkal) dialects. In the 1950s, Vach and Surgut dialects were added. Nothing has ever been published in the Southern Hanty dialects (Dem'an or Konda). Much earlier, in the 1840s, some of the Hanty folklore texts were written down by the Hungarian A. Reguly and Finn M.A. Castren. In the second half of the nineteenth century, some parts of the Holy Bible were translated by the Russian missionaries, who tried to christen the Hanty people. Since 1937, the Hanty script is Russian (Cyrillic). All the dialects mentioned here have plenty of subdialects, which can be called separate languages and whose names cannot be mentioned due to space constraints. Even the main branches are so many that none of them are mentioned. The Northern Hanty dialects (Kazym, Nizjam, Sherkal, Ber'ozovo) have a rather simple system of phonemes, if compared to the southern or eastern dialects. They have only three cases and a rather simple verb system. They have double negation. The eastern dialects (Vach, Vas'ugan, Surgut, Salym, etc.) have vowel harmony, i.e. rigid opposition between front and back vowels. They also preserve the other archaic traits, e.g. a more complicated phonemic system, paradigmatic vowel change, plenty of cases, and semiergative construction of the sentence. The southern dialects are intermediate between the northern and eastern dialects in their phonetical and grammatical characteristics. The ergative construction (as well as in the Northern dialects) is much more seldom, than in the Eastern dialects. Let us consider the similarity of the phonetical systems of two Hanty subdialects to those of Mansi. We took up Kazym (Northern) and Vach-Vas'ugan (Eastern) dialects. The vowel system of the Kazym subdialect of Hanty is very similar to that of the Northern (Sos'va) subdialect of Mansi. It has nine (long and short) vowels [i, I, e:, o, o:, u, u:, a, a:]. According to the work of the active organ of speech (place of articulation), the following four consonantal groups are defined: labio-labial [p, w, m] (no labio-dentals); front [t, s, ?, n, l, ?, r];

palatal [t', s', n', ?', j]; and velar [k, x, ?] and according to the manner of articulation: sonorant [m, w, n, n', l, ?, ?', r, j, ?], occlusive [p, t, t', k], (no affricates), and fricative [s, s', ?, x]. There are no nonsonorant voiceless consonants in the Kazym subdialect of Hanty. The vowel and consonant phonemic system of the Vach-Vas'ugan dialect are more similar to Hungarian than the rest. The following 14 vowels are defined [i, I, e, o, o, o, o, u, u, u, a, a, ?, a], and according to the work of the active organ of speech: labio-labial [p, w, m], (no labio-dental), front [t, s, ?, t?, n, l, ?, l, r], palatal [t', n', l', l', j], and velar [k, q, ?, ?]. According to the manner of articulation, the following can be defined: sonorant [m, n, n', l, l', ?, j, ?], occlusive [p, t, t', k, q], affricate [t?], fricative [w, s, ?, l, l', ?], and voiced consonants [w, l, l', ?]. Each of the described consonantal groups has its influence on the sound structure of the languages speech chain. The high frequency of sonorants may give a melodic picture of a language, while the high frequency of occlusive consonants gives a sharp and abrupt picture of it. Comparing the consonantal sound pictures of these four Ob'-Ugrian languages with the help of 'chi-squared criterion', one comes to the conclusion that the Northern (Sos'va) dialects of Mansi and Northern (Kazym) dialect of Hanty resemble each other most in terms of the distribution of certain types of consonants in their sound chains (75). The sound picture distance between Sos'va and Konda dialects of Mansi is greater (409). This distance between Kazym and the Eastern dialect of Hanty is even greater (476). The distance between Konda dialect of Mansi and Eastern dialect of Hanty is 832. Hungarian is much closer to the Eastern dialect of Hanty (151) than to the Kazym dialect of Hanty (1612) or both Mansi dialects: Sos'va (1018) or Konda (1115). One can see that these distances are much greater than the similarity of Hungarian to Mordovian (174) or Mari (226), the Finno-Ugric of the Volga region.

References

Balandin A. N., Vahrusheva M.P. 1957. The Mansi language, 275. Leningrad: Prosveshch. [in Russian].

Tereshkin N. I. 1966. The Hanty language The languages of the peoples of the USSR, Vol. 3. Finno-Ugric and Samodij languages, 319–42. Moskva: Nauka [in Russian].

Yuri Tambovtsev. 1998. Some phonological features as measure of closeness of dialects. Philologia Fenno-Ugrica 4. 1–19.

Peter Hajdu. 1985. The Uralic Languages and Peoples. Moskva: Nauka [in Russian].

The Hungarian language. Budapest, 1972.

K. E. Majtinskaja. 1993. The Hungarian language Languages of the World. Uralic languages, 179–256. Moskva: Nauka [in Russian].

J. Tompa. 1968. Ungarische Grammatik. Budapest.

YURI TAMBOVTSEV

Hymes, Dell Hathaway

Dell H. Hymes' life and work revolve around an interest in language, a commitment to social justice, and an abiding love of folklore. The intertwining of these themes provided the fertile ground for his views: that language is embedded in culture; that the division between language and culture is a disciplinary artifact, its maintenance a result of protected disciplinary boundaries; and that the only legitimate study of language is socially constituted linguistics. Much of his writing therefore challenges current–traditional practices in anthropology, sociolinguistics, sociology, and formal linguistics. His work reflects a critical and evolutionary Marxist view of knowledge, interpreting the current dialectic as a teleological stage in the development of knowledge, leading, '... to a humanism which can deal with concrete situations, with the inequalities that actually obtain, and help to transform them through knowledge of the ways in which language is actually organized as a human problem and resource' (1996: 60).

Hymes considers the essence of humankind not as an abstraction in the individual, but an actuality entailed in specific social relations. He thus rejects giving priority to an abstract category of language or grammar, and the notion of nativism of language, both positions putting him in opposition to structuralism and Chomskyan linguistics. He met the Chomskyan challenge head on with his formulation of 'communicative competence', which deliberately creates an irony. For Noam Chomsky, competence inheres in the individual, and linguistic rules are verified by the intuitions of the ideal speaker/hearer. The idea that 'communicative' can be collocated with 'competence' turns Chomsky's notion and resulting linguistic discourse on its head. He also refutes the basic premise in linguistics that all languages are equal. Hymes claims that all languages may have equal potential, but because human beings ascribe differing attitudes to different languages and varieties, speakers of some languages and varieties have reduced opportunities.

His strongly held view that forms and meanings are part of social life has influenced both his theoretical writings and his life choices. He is probably most widely known for his formulation of the 'ethnography of speaking', first presented at a meeting of the Anthropological Society of Washington in 1962, the seed for which was germinated in a 1952 debate about the role of linguistics in anthropology, one that drew scholars from several disciplines. That ethnography would become his approach to language in society was rooted in his mixed training and later interactions with anthropologists, linguists and social scientists, at Harvard, and at Berkeley. He chose the term linguistic anthropology, making it a branch of anthropology. This approach resonates with his view of language's role in social life and by applying it to the examination of language, challenged both linguistics and anthropology. Using ethnography, anthropologists uncover the rules and symbols that hold meaning for the community being studied. Such an approach was new to linguistics, which instead considered descriptions of phonology and grammar as the principal framework. By choosing the term 'speaking', he also challenged anthropologists who, when studying cultures of a community, took speech for granted.

This paper launched a new perspective, providing the parameters for analyzing and describing speech communities—grouping members who share one or more of its 'ways of speaking'. Members who know not only the rules of speaking but also the sociocultural norms and values that guide interaction and cultural behavior within specific communities can be considered to have communicative competence. His framework for description included speech community, speech event (activities governed by rules for speaking), factors in a speech event (e.g. sender), and the functions of a speech event (e.g. persuasive).

Over the next decade, the framework was revisited as colleagues and students conducted ethnographies in different settings—nonwestern, Native American, and, increasingly, schools. In 1972, he elaborated on the factors in the speech event, devising the mnemonic SPEAKING: setting, participants, ends, act characteristics, key (tone), instrumentalities, norms, and genres. This list has been compared to Michael Halliday's field, tenor, mode, and was found to be unmotivated linguistically. Hymes, however, only ever meant SPEAKING to be a heuristic, a list of questions to ask when examining language in context. For both, the bundling of linguistic features in a community results in linguistic repertoires available to that community. Ethnography of speaking was broadened to the 'ethnography of communication', Hymes never intended to ignore written language by his use of 'speaking'; rather, he used it as a surrogate for all communication modes, including gestures and, most recently, electronic modes.

Hymes was the driving force behind a new journal, calling it *Language in Society*, not a journal of sociolinguistics, so that its concern would focus on

social/cultural life. Perhaps, his greatest source of influence was his 20 years of editing the journal, many of the articles owing much to his imaginative thinking as he dialogued with the author, sending detailed recommendations, comments, and further ideas. At the University of Pennsylvania, he moved from Anthropology to Folklore and Linguistics, because of political division within the department, even though he felt that anthropology was in a unique position for studying the sociology of language from a comparative perspective. Later still, he moved to the School of Education as Dean, a move that was not just practical (to save the School from closure) but also philosophical. He had already set a new direction for educational research with his volume coedited with Cazden and John (1972), by insisting on the need to examine actual classroom interaction in order to understand the role of language in teaching and learning. In education, his beliefs that understanding how language varies and what that variation means for speakers, especially in terms of inequality, influenced both theory and practice leading to the development of a program in educational linguistics. Education was also a platform for advocating that all human beings could benefit from an understanding of language. His most recent work has focused on analyzing oral narratives, bringing together his love of folklore, his commitment to social justice, and his insistence that language should always be examined in its social context.

Biography

Dell Hathaway Hymes was born in Portland, Oregon on June 7, 1927. He received his B.A. in 1950 from Reed College; M.A. in 1953 from Indiana University; and Ph.D. in 1955 for a dissertation that was a linguistic analysis of Kathlamet Chinook Indian texts that had been collected by Franz Boas, Indiana University. He became Instructor, then Assistant Professor, Social Relations Department, Harvard University in, 1955–1960. Also, he became Associate Professor, then Professor, Department of Anthropology, University of California, Berkeley in 1960–1965; he became Professor, Anthropology, University of Pennsylvania in 1965–1972; Professor, Folklore and Linguistics, University of Pennsylvania in 1972–1988; Professor, Sociology, University of Pennsylvania in 1974–1988; Dean of the School of Education, University of Pennsylvania in 1975–1987; Professor of Anthropology and English, University of Virginia in 1987–1990; and Commonwealth Professor of Anthropology, University

of Virginia in 1990–1998, now Emeritus. He was also Fellow of the American Association for the Advancement of Science; Life Fellow, Clare Hall, Cambridge; member of the British Academy; Fellow at the Center for Advanced Study in the Behavioral Sciences, 1957–1958; Guggenheim Fellow, 1969–1970; Senior Fellow, National Endowment for the Humanities, 1972–1973; Trustee, Center for Applied Linguistics, 1973–1978; President, American Folklore Society, 1973–1974; President, Council on Anthropology and Education, 1978; President, Council on Anthropology and Education, 1978; President, Linguistic Society of America, 1982; President, American Anthropological Association, 1983; and President, American Association for Applied Linguistics, 1986. He was also founding coeditor of *Language In Society* with William Labov and Allen Grimshaw, 1972–1992.

References

Bauman, Richard, and Joel Sherzer (eds.). 1982. Explorations in the ethnography of speaking. Cambridge: Cambridge University Press.

Heath, Shirley Brice. 1983. Ways with words. Cambridge: Cambridge University Press.

Hornberger, Nancy. 1988. Bilingual education and language maintenance: a Southern Peruvian Quechua case. Berlin: Mouton.

Hymes, Dell Hathaway. 1969. Reinventing Anthropology.

———. 1972. The scope of sociolinguistics. Sociolinguistics: current trends and prospects. 23rd annual Georgetown roundtable, ed. by Roget W. Shuy. Washington, DC: Georgetown University Press.

———. 1986[1972]. Directions in sociolinguistics: the ethnography of communication, ed. by John Gumperz and Dell Hymes. New York: Basil Blackwell.

———. 1972. Functions of language in the classroom, ed. by Courtney B. Cazden, Vera P. John, and Dell H. Hymes, New York: Teachers College Press.

———. 1974. Ways of speaking. In Bauman and Sherzer.

———. 1974. Foundations in sociolinguistics: an ethnographic approach.

———. 1983. Studies in the history of linguistic anthropology.

———. 1996. Ethnography, linguistics, narrative inequality: toward an understanding of voice.

McDermott, Ron. 1977. The ethnography of speaking and reading: Linguistic theory: what can it say about reading, ed. by Roger Shuy. Newark, Delaware: International Reading Association.

Philips, Susan Urmston. 1983. The invisible culture: communication in classroom and community on the Warm Springs Indian Reservation. New York: Longman.

Saville-Troike, Muriel. 1989. The ethnography of communication. New York: Basil Blackwell.

DENISE MURRAY

Iconicity

Many believe that language is arbitrary. Arbitrariness refers to a random relation between a word and its meaning. However, there are many examples of nonarbitrariness in language known as iconicity. Iconicity is the opposite of arbitrariness because iconicity refers to a resembling relation between a word and its meaning. A word is directly related to its meaning in many ways.

The form of words like *cockatoo, cuckoo, buzz, thump, crash, splash, thrash, clash, hiss, puff,* and *purr* is connected to their meanings. These words imitate the sounds created by the referents. While, '*ha, ha, ha*', captures the sarcastic laughter in English, '*he, he, he*' is commonly used after a remark intended as a joke in Malay e-mails as well as Singapore and Malaysian English e-mails. These words do not come into language by chance. The derivation of these words is based on certain comprehension based on sound. The words are sound images of the various meanings. This type of iconicity is termed imagic iconicity.

Imagic iconicity comes in the form of sound symbolism. Sound symbolism occurs if a particular sound informs a particular sense across various words. Examples in Malay, Cantonese, English, and Spanish words suggest that the vowel [o] symbolizes a negative sense across languages (Table 1).

Current bad English o-iconic words of the world are *bomb, romp,* and *storm* all of which had *brought down* lives in all four corners of the world resulting in either death or disgrace.

Another type of nonarbitrariness is diagrammatic iconicity. Diagrammatic iconicity suggests that a word is a diagram of a meaning. Linguistic signs are used to illustrate some structures of the meaning. Sign language is a form of meaning depiction. The notions of *up* and *down* in various sign languages relate closely to the vertical directions of these concepts. The resemblance between a raptor snatching a sitting-up rodent and one hand snatching at an upright finger in American Sign Language is a good example of diagrammatic-iconic gesture. However, believers of language signs must arbitrarily deny sign language as language (see Armstrong, Stokoe, and Wilcox).

Motivation is the operating force behind diagrammatic iconicity. Motivation in iconicity refers to perceived similarity between the structure of the expression and the concept represented in the expression. The need to name things results in the expansion of meaning motivated by resemblance in appearance. A schematic resemblance between a tall *crane* machine and the neck of *crane* motivates the derivation of a new meaning in the word *crane*. The creation of new words may be motivated by functional similarity, e.g. the shuttling of a *shuttlecock* and a *space shuttle* in the air.

Quantity iconicity is another type of diagrammatic iconicity. Quantity iconicity holds the basic notion that more form entails more meaning. Physical increase in quantity is captured loosely by more words in reduplication. Hence, plurals usually have more words. Simple noun reduplication in Malay exemplifies quantity iconicity (Table 2).

Diagrammatic iconicity records the flow of time from its commencement until its completion. This sequence of time and the natural order of an event unfolding in time are reflected in sentences in sequential iconicity, e.g. '*Syukri went to school on Monday*

TABLE 1

Malay	Cantonese	English	Spanish
hodoh (ugly)	*soh* (slow in thinking)	*horror*	*tonto* (stupid)
bodoh (stupid)	*lõw* (man in a rude sense)	*sorrow*	*zonzo* (stupid)
kotor (dirty)	*põr* (woman in a rude sense)	*fool*	*coco* (funny looking)
rogol (rape)	*mór* (devil)	*loon*	*memo* (silly, foolish)
bocor (leak)	*mõng* (dumb)	*moron*	*chato* (flat nose)

TABLE 2

Basic Singular Noun in Malay	Basic Plural Noun in Malay
pelajar (a student)	*pelajar-pelajar* (students)
sekolah (a school)	*sekolah-sekolah* (schools)
gunung (a mountain)	*gunung-ganang* (ranges of mountain)
sayur (a vegetable)	*sayur-mayur* (vegetables of all sort)

and left for Singapore on Tuesday in the same week,' rather than '**Syukri left for Singapore on Tuesday and went to school on Monday in the same week.'* In fact, sequential iconicity is a refection of common sense, which is imperative to sensible sentences like:

> *Jyh cooked the fish with ketchup before he ate it.*
> **Jyh ate the fish with ketchup before he cooked it.*
> *Siti knocked the vase, and consequently it broke.*
> **Siti broke the vase and consequently it knocked.*

The language of distance is iconic. The vivid sense of *near* is easier to register in the mind than the distant notion of *far*. These two senses are codified in phrases like *now and then, this and that,* and *in and out*. That *near* comes before *far* in language, just like their real-world relation illustrates syntactic iconicity. Syntactic iconicity refers to the fixed order in phrases. These phrases are known as freezes. Syntactic iconicity in *here and there* has equivalents like *ingum angum* in Tamil, *sini sana* in Malay, *li do guo do* in Cantonese, *ci dau he dau* in Hokkien, *ci gou heo gou* in Teochew, and *zhe li na li* in Mandarin.

Likewise, *above* precedes *below*, *on* precedes *under*, and *up* precedes *down* (see Landsberg). Freezes like *ups and downs, head to toe, head over heel, top-down, high and low* are common examples. Quantity decrease in freezes like *more or less*, and *big to small* are found in Mandarin and Malay. Unlike the Malay freezes *lebih kurang* (more or less), the Indonesian counterpart *kurang lebih* (less or more) does not reflect the iconicity of quantity decrease. This is due to the politeness convention that prefers a lesser or weaker meaning before a stronger one.

Freezes like *sir and madam*, *father and mother*, *papa and mama*, *Adam and Eve*, *king and queen*, *boys and girls*, *husband and wife*, *Mr. and Mrs.*, and *brothers and sisters* show that *man* syntactically precedes *woman*. That man comes before woman is based on the perception of man as the center of the universe or the *me-first* principle. This is quite true in the ancient and current real world. Contradictions like *mom and dad*, *bride and groom*, *aunts and uncles*, and *ladies and gentlemen* can be explained either by the politeness convention, the stronger phonological laws, whereby high vowels precede low vowels or the less-syllabic items precede more-syllabic items order (see Cooper and Ross).

References

Armstrong, David F., William C. Stokoe, and Sherman E. Wilcox. 1996. Gesture and the nature of language. Cambridge: Cambridge University Press.

Askedal, John O. 1995. Iconicity in the basic serialization rules of Modern German. In Landsberg.

Cooper, William E., and John R. Ross. 1975. World Order. Papers from the parasession on functionalism of the Chicago Linguistic Society.

Haiman, John. 1985. Natural syntax: iconicity in language. Cambridge: Cambridge University Press.

Landsberg, Marge E. (ed.) 1995. Syntactic iconicity and linguistic freezes: the human dimension. Berlin: Mouton de Gruyter.

Landsberg, Marge E. 1995. Semantic constraints on phonologically independent freezes. In Landsberg.

Malkiel, Yakov. 1990. Diachronic problems in phonosymbolism: Edita and Inedita. Amsterdam: Benjamins.

Sew, Jyh Wee. 1995. Sound meanings in Malay. M.A. Thesis. The National University of Singapore.

Sew, Jyh Wee. 1998. Review of Marge E. Landsberg ed., Syntactic iconicity and linguistic freezes: the human dimension. WORD 49.

Sew, Jyh Wee. 1998. Review of Friedrich Ungerer and Hans-Jörg Schmid: introduction to cognitive linguistics. The Web Journal of Modern Language and Linguistics 3.

http://www.ncl.ac.uk/~njw5/issue03/jyh-wee.htm

JYH WEE SEW

See also **Onomatopoeia**

Identity and Language

Throughout human history, language and identity have been closely linked. Political boundaries have been drawn along linguistic lines, establishing nationality as well as ethnic division. Often, a single feature of pronunciation can provide a convenient means of identification. Pronunciation, grammatical and lexical markers, naming, and discourses may all be used to establish identity. The linguistic approach that is most centrally concerned with the interrelationship between identity and language is sociolinguistics. Sociolinguistic approaches to identity and language have changed over time. At least three different strands of inquiry into identity and language can be identified: variationist, interactional, and social constructionist approaches.

Variationist sociolinguistics, which is strongly associated with the pioneering work of William Labov, is based on the premise that language reflects identity. Assuming that a person's speech reflects characteristics of that person, aspects of social identity—such as social class, sex, ethnicity, nationality, or age—were traced through forms of language. The bulk of variationist work on identity and language concerned pronunciation. Typically, certain pronunciation variables were linked to aspects of social identity. However, in recent years, work in variationist sociolinguistics has been challenged on the grounds that social identity cannot provide an explanation for language use because it is a concept that is itself in need of explanation. Critics of variationist sociolinguistics argue that instead of asking, 'How do women and men speak differently?' or 'How do Black and White Americans speak differently?' it would be more important to find out how people become gendered or racialized beings in the first place.

Interactional sociolinguistic approaches to identity and language, as pioneered by John Gumperz and used by Deborah Tannen, draw largely on social psychology. In this approach, identity is based on group membership, where the negotiation of identity in interaction is of central concern. Instead of focusing on pronunciation, interactional sociolinguists study discourse, written or spoken communication, preferring qualitative methods over quantitative in data collection and analysis. Interactional sociolinguistics assumes that identity as group-membership is negotiated, challenged, or upheld in conversation. When people come together from different backgrounds, misunderstandings are likely to occur, as they do not share the same conventions for signaling identity and role relationships. Consequently, most of the work in this tradition has been concerned with cross-cultural communication. 'Cross-cultural' has been rather loosely defined and includes interactions between people from different national backgrounds (e.g. American–Japanese business communication), between migrants and natives (e.g. South-East Asians in British job interviews), between people from different ethnic backgrounds (e.g. Black and White Americans), or between women and men. This social psychological conception of identity, however, has come under attack for at least two different reasons. In this approach, identity as group membership is seen as essentializing, meaning identity is regarded as internal and immutable. Further, this view of identity relies on the idea of a homogeneous group while, in fact, people are members of many different groups simultaneously (e.g. American, Hispanic, professor, female, heterosexual). The latter objection, that social identity is hybrid and heterogeneous, is raised by linguists who work in bi- and multilingualism studies. The language use of bilinguals clearly exemplifies that speakers do not only have one single identity but rather a repertoire of identities.

Contemporary approaches to language and identity most often build on the framework of social constructionism. Identity is now seen as relational, cultural, and contingent. It is relational because it is located in connections made between people rather than in the minds of individuals. Identity is cultural because it is based on shared understandings, and it is contingent because it is a strategic performance that may fail or misfire. Speakers are seen as strategically deploying their linguistic repertoires in order to project chosen identities. However, their acts of identity are not exclusively in their own hands as they depend (a) on the repertoires at their disposal and (b) on the ways in which their speech partners choose to view them. Identities are constructed or coconstructed depending on the power relationship that pertains between the interactants. In many societies, categories such as caste, gender, or race are imposed and coerced, leaving little or no room for individuals to perform or explore alternative identities that might deviate from prevailing ideologies. Three aspects of social identity that have played a central role in recent linguistic work will be discussed based on the following: gender, nationality, and bilingualism.

Within feminist and gender theory, identity has become the central question since, in the 1980s and 1990s, women of color and lesbian theorists started to challenge the assumption that the experience of white

heterosexual middle-class American women was universal. Consequently, language and gender researchers have begun to explore the linguistic construction of fluid, hybrid, and alternative identities. Recent research (see Bucholtz et al., 1999) explores how identity is invented (e.g. through the creation of shared experience in the coming-out stories of deaf and hearing lesbians) and how the performance and perception of identity is constrained through belief systems (e.g. through the 'woman as dessert' metaphor in English). Further foci include speakers' individual creative responses to cultural ideologies and identity constructions as 'on-line' events that may never be complete (e.g. callers on a shopping channel may position themselves simultaneously as consumers with questions about the product and experts who are familiar with consumption patterns).

Unlike gender, nationality has been seen as intricately linked to language for centuries. To the present day, many nations, particularly in Europe and North America, subscribe to a 'one nation, one language' ideology. Many states have used the reification of a standard language and the common identity that supposedly goes with a shared language as a means to naturalize political borders. In an intricate genre analysis, Wodak et al. (1999) describe how 'the same' national identity (Austrian) is performed in significantly different ways and with significantly different meanings in public and private speech (e.g. newspaper articles vs. one-to-one interviews).

For some time, researchers into multilingualism have seen language choice as meaningful and as an 'act of identity' (see Le Page and Tabouret-Keller 1985), whereby speakers proclaim their allegiance to a particular speech community. Code-switching is often seen as a means to signal changes in the role relationship that pertains between interactants. Furthermore, second-language acquisition theory has often viewed successful language learners as good assimilators. The easier they find it to shed their native identity and assimilate to the one related to the target language, the more successful their learning outcomes were thought to be. Only recently, linguists (e.g. Pavlenko et al., 2001) have come to acknowledge that bilinguals and second-language learners tend to engage in more than one community and that they may have different identities in different languages.

At present, the study of identity and language is a lively and diverse field of research, encompassing a wide array of issues and implications. Given the overall global climate of identity politics, the scope of sociolinguistic study, both theoretical and practical, is likely to increase in the future.

References

Anderson, Benedict. 1983. Imagined communities: reflections on the origin and spread of nationalism. London: Verso; 2nd edition, 1991.

Bucholtz, Mary, A.C. Liang, and Laurel A. Sutton (eds.) 1999. Reinventing identities: the gendered self in discourse. New York and Oxford: Oxford University Press.

Eckert, Penelope. 2000. Language variation as social practice: the linguistic construction of identity in Belten High. Oxford and Cambridge, MA: Blackwell.

Gumperz, John J. (ed.) 1982. Language and social identity. Cambridge: Cambridge University Press.

Heller, Monica. 1999. Linguistic minorities and modernity: a sociolinguistic ethnography. London: Longman.

Le Page, Robert, and Andrée Tabouret-Keller. 1985. Acts of identity: Creole-based approaches to language and ethnicity. Cambridge: Cambridge University Press.

Mühlhäusler, Peter, and Rom Harré. 1990. Pronouns and people: the linguistic construction of social and personal identity. Oxford: Basil Blackwell.

Norton, Bonny P. 2000. Identity and language learning: gender, ethnicity and educational change. London: Longman.

Pavlenko, Aneta, and Adrian Blackledge (eds.) 2001. Negotiation of identities in multilingual contexts. (Special issue of The International Journal of Bilingualism 5).

Pavlenko, Aneta, Adrian Blackledge, Ingrid Piller, and Marya Teutsch-Dwyer (ed.) 2001. Multilingualism, second language learning, and gender. Berlin and New York: Mouton de Gruyter.

Piller, Ingrid. 2002. Bilingual couples talk: the discursive construction of hybridity. Amsterdam: Benjamins.

Shotter, John, and Kenneth J. Gergen (eds.) 1989. Texts of identity. London, Newbury Park and New Delhi: Sage.

Wodak, Ruth, Rudolf de Cilia, Martin Reisigl, and Karin Liebhart. 1999. The discursive construction of national identity. Edinburgh: Edinburgh University Press.

INGRID PILLER

Ideology and Language

Ideology's features are topics of complex debate. Often defined as matrices of ideas and images inhabiting individuals' minds, ideologies are also thought of as neither ideas nor images, but as practices that are embedded in and reproductive of specific socioeconomic structures. Ideology can empower or oppress. Furthermore, ideology is related to language in numerous ways.

An ideology may empower persons when it defines their ethnic, gender, class, or other identity, organizes a 'worldview' that gives meaning and coherence to their lives, or facilitates their entrance into social, civic, or political life. Oppressive ideologies marginalize, stigmatize, or otherwise define persons to reproduce their exploitation or subordination in a socioeconomic hierarchy.

A focus on an ideology as empowering may downplay the relations of ideology and knowledge. Sociologists comparing various cultures' ideologies may want to know the differences among ideologies rather than to judge their compliance to a normative definition of knowledge. A barrier to this unbiased evaluation is the sociologists' own assumptions about knowledge. The 'critique of ideology' exposes such mystifying errors or beliefs by which an ideology reinforces structures of oppression. Often the same ideology empowers some, oppresses others, and gains hegemony. An ideology becomes hegemonic by defining the arguments contending groups use to debate with each other; by prescribing the values to which clashing ideas must adhere to achieve legitimacy; or by projecting a totality in which diverse political actors imagine an open network of cultural, social, and economic relations to be enclosed.

Debunking an oppressive ideology's knowledge claims contests that ideology. Theorists have long debated whether ideology can be distinguished from knowledge. The European Enlightenment's attack on religious dogma claimed that ideologies are illusions that reason can dispel to foster an ideology-free society. Karl Marx defined ideology as mystification in the interest of a ruling class. Like Marx, the twentieth-century communist philosopher Louis Althusser claims that scientific knowledge is distinct from ideology. Because of ideology's role in subject formation, Althusser concludes that ideology will always exist, even in a future classless society. Contemporary 'post-Marxist' theories include ethnicity, gender, and culture as important determinants of ideology. Questioning the opposition between science and ideology, post-Marxists doubt the theorists' ability to contest oppressive ideologies effectively. Post-Marxists define political struggle as a battle for hegemony in which no participant holds the cognitive or ideological high ground.

Ideology intersects language when politics privileges one language and denigrates, restricts, or excludes others. California voters have passed a series of proposals, making English the state's 'official language', requiring that government agencies use English to the exclusion of other languages, and outlawing bilingual instruction in the state's schools, exemplifying ideology's impact on language.

Californians write and speak numerous languages, but enough Californians felt threatened by this language diversity to institutionalize an ideological monolingualism that values social unity as a function of linguistic homogeneity. The propositions' success arguably related to anxieties concerning California's growing population of Spanish speakers. These examples raise the issue of 'linguistic racism', in which prejudicial attitudes about an ethnic group extend to that group's language. Such racism often assumes that one language is inherently more beautiful or a better vehicle for thought than others. And, in the form of slurs, jokes, and ideologically coded diction, language can influence racial ideology.

Language also reflects ideologies of sexuality and gender. For example, consider the following sentence: 'A linguist, if he wants a successful career, masters contemporary linguistic research on English'. This sentence, by assuming that the typical linguist is male, betrays sexist beliefs concerning women's abilities and roles in society. Sexism in language use not only signals an ideology; sexism as a relation of power embedded in institutions reproduces itself through such uses when they are uncontested. Similarly, racism in language use reveals an ideology and contributes to the continuation of racist power relations.

Even in the study of language, ideologies present barriers to unbiased evaluation. Our sentence also assumes that English should be privileged in linguistic inquiry. The phrase 'language ideology' refers both to ideologies found in language and to ideologies about language. When power relations dictate which languages may be studied or what constitutes their legitimate study, we confront language ideology's influence on the production of linguistic knowledge.

The study of language itself reflects a language ideology that assumes languages are static entities unaffected by their ideological contexts. These contexts sometimes influence a language's pattern of sound change or even precipitate its demise. Examples of demise are the fates many indigenous peoples' languages have suffered in the wake of Europe's colonization of Asia, Africa, Australia, and the 'New World', direct victims of racist ideology. The recent trend to install Standard American English as a global lingua franca endangers ever more infrequently used languages. Language reflects ideologies, oppressive and progressive, which in turn influence the evolution of those same languages. Further, this reciprocal relationship also holds in the political arena. Language not only influences ideologies but may also be used as a tool to control groups or thought. The study of the intersection of language and ideology contributes to our understanding of the political implications of global socioeconomic realities.

491

References

Althusser, Louis. 1970. Idéologie et appareils idéologiques d'état (Notes pour une recherche). La Pensée 151; as Ideology and ideological state apparatuses: notes towards an investigation. Lenin and philosophy, translated by Ben Brewster. London: New Left Books, and New York: Monthly Review Press, 1971.

Barrett, Michèle. 1991. The politics of truth: from Marx to Foucault. Cambridge: Polity Press, and Stanford: Stanford University Press.

Calvet, Louis-Jean. 1987. La guerre des langues et les politiques linguistiques. Paris: Payot; as Language wars and linguistic politics, translated by Michel Petheram, Oxford and New York: Oxford University Press, 1998.

Chomsky, Noam. 1989. Necessary illusions: thought control in democratic societies. London: Pluto, and Boston: South End Press.

Derrida, Jacques. 1996. Le monolinguisme de l'autre: ou la prothèse d'origine, Paris: Editions Galilée; as Monolingualism of the other; or, the prosthesis of origin, translated by Patrick Mensah. Stanford: Stanford University Press, 1998.

Eagleton, Terry. 1991. Ideology: an introduction. London and New York: Verso.

Edgley, Alison. 2000. The social and political thought of Noam Chomsky. London and New York: Routledge.

Joseph, John, E., and Talbot, J. Taylor (eds.) 1990. Ideologies of language. London and New York: Routledge.

Kress, Gunther, and Robert Hodge. 1979. Language as ideology, London and Boston: Routledge & Kegan Paul; 2nd edition, London and New York: Routledge, 1992.

Nettle, Daniel, and Suzanne Romaine. 2000. Vanishing voices: the extinction of the World's languages. Oxford and New York: Oxford University Press.

Pêcheux, Michel. 1975. Les vérités de la Palice. Paris: F. Maspero; as Language, semantics, and ideology, translated by Harbans Nagpal, London: Macmillan Press, and New York: St. Martin's Press, 1982.

Woolard, Kathryn, A. 1998. Language ideology as a field of inquiry. Language ideologies: practice and theory, ed. by Bambi, B. Schieffelin, Kathryn, A. Woolard, and Paul, V. Kroskrity. New York and Oxford: Oxford University Press.

ROBERT S. OVENTILE

Idiomaticity

The traditional and most widespread definition of idiomaticity maintains that it is the result of a mismatch between meaning and form: the overall meaning of an idiomatic expression is figurative and noncompositional, that is, it cannot be deduced from the literal meaning of its individual components. Idiomatic expressions are thus words, phrases, or even whole sentences that seem to function as an unanalyzable unit: *egghead, to bury the hatchet, the early bird catches the worm,* and the like.

The study of idiomaticity has advanced greatly since the 1980s because of a change in the discipline of linguistics that was characterized by an increasing interest in meaning and discourse. This universal grammatical phenomenon has begun to receive the attention it did not receive during the long period when sentence structure was the principal focus of linguistic investigation.

Although unpredictable from its form, the meaning of an idiomatic expression cannot be considered purely random; it is the direct consequence of a historical process of extension, whereby the literal, compositional meaning of that pattern, although not completely lost, is gradually replaced by a new metaphorical, noncompositional meaning that becomes widely accepted in society. Contrary to what might be expected, however, the connection between the two senses is not easily recognizable at the present time in every linguistic construction that counts as an idiom. Hence, there is a distinction between 'transparent' idiomatic expressions like *skate on thin ice,* which have a clear literal origin, and 'opaque' ones like *to pull someone's leg,* which do not display any logical relation with their source. Idiomaticity is thus the outcome of a historical process that must nonetheless be analyzed as a rather static phenomenon.

But classic theories of idiomaticity reject this distinction because, according to them, all idiomatic expressions are dead metaphors that lack any connection with their original sources. As a consequence, the relation between their form and meaning is traditionally said to be arbitrary. Recent research within the theoretical framework called 'cognitive grammar' has shown, by contrast, that the form–meaning relation underlying many idiomatic expressions is far from arbitrary—that it is motivated not only by their primitive meaning but also by conventional images and conceptual metaphors. Thus, the traditional analysis asserts that nothing explicit in *to spill the beans,* for instance, justifies its meaning 'to reveal a secret'; this new approach, however, claims that such a figurative meaning is clearly determined by the 'conduit metaphor'—that is, a metaphor that compares the mind with a container and ideas with physical entities.

Apart from shedding some light on the semantic motivation underlying idiomaticity, the afore-mentioned historical process is also noteworthy for two further reasons: (a) it is responsible for the institutionalized meaning of idiomatic expressions; (b) it explains their frequent coexistence with literal interpretations of the same expressions.

(a) Along with the nonliteralness and non-compositionality of an idiomatic expression, its sense also has to be institutionalized; that is, it must be regularly associated by the members of a speech community with a specific form. The number of people accepting such form–meaning asymmetry is considered essential in this slow process of adaptation. They can range from a small in-group to an entire linguistic community, hence, their number influences the degree of institutionalization of a specific idiomatic pattern and consequently its degree of idiomaticity: that is, the more widely used in society an idiomatic expression is, the more institutionalized and idiomatic it is, hence the contrast between those idiomatic constructions that are listed in dictionaries as approved by the entire society and those others, like slang, jargon, and familiar euphemisms that have a private meaning accessible only to small social groups. Regardless, during the institutionalization process they all acquire a national and cultural flavor that makes understanding and acquiring them highly difficult for the outsider.

(b) The institutionalization process is also significant because, although it does not imply the systematic loss of the original construction that serves as input for the idiomatic pattern, it explains the frequent coexistence of many idiomatic expressions with literal interpretations of the same expressions. The commonly suggested hypothesis that idioms are usually ambiguous between a metaphorical and a literal interpretation is thus reinforced. However, such ambiguity can no longer be considered a distinguishing property of idiomaticity, due to the following results of recent investigations based on immense collections of actual language data: (a) a vast number of idiomatic patterns, such as *rain cats and dogs* and *by and large,* lack literal interpretations, either because they would violate truth conditions or because they would be ungrammatical, (b) the literal interpretation of whatever idiomatic expressions do allow both readings is highly infrequent in actual language data, because the discourse where they occur generally favors the figurative meaning over the literal one. Even when no discourse element serves this clarifying purpose, the alleged ambiguity can still be removed via pronunciation: the idiomatic reading, in accordance with its form–meaning unity, requires shorter pauses and word durations than its literal counterpart.

Moreover, new studies have shown that idiomatic expressions are less exceptional than traditionally thought: (a) Far from being structurally frozen patterns, they present more structural variety than has been asserted. Specifically, the more semantically motivated and institutionalized the meaning of an idiomatic expression, the more structural flexibility it shows. (b) The absolute syntactic rigidity that does in fact characterize some idiomatic patterns also appears in a large number of nonidiomatic constructions: *to kick the bucket,* for instance, does not have a passive form, i.e. *the bucket was kicked (by him)* is unacceptable; however, the nonidiomatic construction *she has two sons* does not have a passive either: *two sons are had (by her)* is equally unacceptable. The traditional assumption that special rules apply to idiomatic language must thus be rejected in view of these facts.

In sum, idiomaticity is a heterogeneous and measurable linguistic phenomenon, variable in meaning, form, and discourse functions. It thus forms a continuum with different degrees of idiomaticity that cannot be straightforwardly defined according to one single parameter.

References

Cacciari, Cristina, and Sam Glucksberg. 1991. Understanding idiomatic expressions: the contribution of word meanings. Understanding word and sentence, ed. by G.B. Simpson, 217–40. North-Holland: Elsevier Science.

Chafe, Wallace. 1968. Idiomaticity as an anomaly in the Chomskyan paradigm. Foundations of Language 4. 109–27.

Fernando, Chitra, and Roger Flavell. 1981. On idiom. Critical views and perspectives. Exeter: University of Exeter.

Fraser, Bruce. 1970. Idioms within a transformational grammar. Foundations of Language 6. 22–42.

Gibbs Jr., Raymond W., and Jennifer E. O'Brien. 1990. Idioms and mental imagery: the metaphorical motivation for idiomatic meaning, Cognition 36. 35–68.

Jackendoff, Ray. 1997. The architecture of language faculty, Cambridge, MA: The MIT Press.

Jaeger, Leon. 1999. The nature of idioms. a systematic approach. Berne, Berlin, Brussels, Frankfurt, New York and Vienna: Peter Lang.

Lakoff, George. 1987. Women, fire, and dangerous things. What categories reveal about the mind. Chicago: The University of Chicago Press.

Makkai, Adam. 1971. Idiom structure in English. The Hague: Mouton. 171–87.

Makkai, Adam. 1987. Idiomaticity and phraseology in Psot-Chomskyan linguistics: the coming-of-age of semantics beyond the sentence. Semiotica 64(1/2).

Moon, Rosamund. 1998. Fixed expressions and idioms in English. A corpus-based approach. Oxford: Clarendon Press.

Strässler, Jürg. 1982. Idioms in English. A pragmatic analysis. Tübingen: Gunter Narr Verlag.

BEATRIZ RODRÍGUEZ-ARRIZABALAGA

Idioms

An idiom can be defined as a set phrase, semantically opaque, whose meaning cannot be deduced from the meanings of its constituents. Although it shows an expressive and particular value, and sometimes a (lexical or syntactic) archaic structure, it behaves as a single semantic unit. Expressions such as *a night owl, a tempest in a teapot, not enough room to swing a cat, to leave no stone unturned, to knock on wood, to put the cart before the horse* are called idioms.

A set phrase is a sequence that has experienced fixation, the process by which a multiword expression whose constituents are free (or were free when the phrase was created) becomes an expression whose constituents are integral parts of it. Therefore, the elements lose their own meaning in favor of the global meaning of the expression and the phrase becomes a new lexical unit, independent, with its own new meaning.

Thus, idioms only correspond to one signified—they have a noncomplex, simple meaning. This means that they can be substituted by simple words as they are expressive alternatives to a neutral concept that already exists in the language. For example, *when hell freezes over* can be replaced by *never*. Idioms constitute a luxury of the lexicon in as much as it is possible to communicate basically without them. Therefore, understanding idioms shows that the speaker has a very good knowledge of the language in which he expresses himself (most of the time, idioms are the last aspect that speakers identify when they learn a foreign language).

Because they are fixed phrases, idioms cannot vary in their form, with the exception of some elements that enable them to adapt to the specific context. For example, when the idiom is a verbal unit, the space of the subject is free and the tense of the verb is to be chosen according to the context: Bart/My brother/The captain… *broke the ice*. The possessive can also be adapted: *I will keep* my *fingers crossed* for you.

Idioms are semantically opaque. That is to say that they cannot be understood literally, in a compositional way. They cannot be interpreted from the total of each word that constitutes it. Even if you know and understand each constituent, the meaning of the whole phrase remains confusing. For example, *white* and *lie* are words that are common and easily understood, but put together, they do not have the same meaning. The compound must be learned as if it were a new word. Thus, it is impossible to deduce the global meaning from the total of the signifieds just as it is not possible

to isolate the different matters of a chemical combination.

This can be explained by the fact that the semantic features selected are not the same ones whether the constituents are used when they are free (if they still exist outside of the phrase) or as parts of the idiom. For this reason, most of these utterances can be polysemous—they may have multiple meanings. They can be interpreted in either a literal or a figurative manner. Expressions are exocentric when the total of meanings does not enable the whole phrase to be understood. For example, an expression such as *to be in the doghouse* has two possible interpretations. When the speaker says: "*he is in the doghouse*", the expression is endocentric if the subject is a dog, and exocentric if the subject is a man (it means the man is in trouble with his wife). Thus, the endocentric utterance, whose meaning is compositional, is to be interpreted in a literal way (it is a common utterance). The exocentric statement is fixed and to be interpreted in the figurative way; it is therefore an idiom.

When the phrase is exocentric, the speaker does not conceptualize systematically the meaning of each constituent. For example, in an expression such as *to have a green thumb*, he will not imagine he has a colored finger. For a nonnative speaker (or somebody unfamiliar with the set phrase and its global meaning), his first interpretation will be literal even though the phrase must only be interpreted as a unit. Because they are set phrases, idioms are often used to create puns. One constituent replaces another, or the other meaning of the polysemous word is selected. This unexpected aspect of the new phrase or the new signification causes humor, by reverting back to the literal form. For example, using the idiom *to work for peanuts* can lead to the following play on words: 'Charlie Brown has spent his entire life working for Peanuts.'

Thus, in idioms, words are used figuratively, through two main figures of speech: metaphor and metonymy. This matter of fact gives life and richness to the language by enabling it to absorb new concepts that need to be expressed linguistically in a new way. The use of metaphors is motivated by the incapacity of the human mind to reach a complete state of abstraction. To make an abstract notion accessible and to give it more weight, the speaker compares it with a physical object he knows. He uses basic schemes in order to conceptualize his experience and the external world (even if the analogy appears a little illogical): *to bear*

one's cross, to have a monkey on one's back, to have a frog in one's throat. Thus, metaphors are based on similarity: metaphoric idioms appear in their simpler form when they are built around the adverbs 'as' or 'like': *as innocent as a lamb, as solid as a rock, like (two) peas in a pod, like a bolt out of the blue*. For its part, metonymy is based on contiguity: one word or phrase is substituted for another with which it is closely associated. But it is also an erroneous perception, which consists of confusing two different things, for example: *to put bread on the table* (bread for money), *to give a hand* (hand for help), *heads will roll* (heads for people).

Sometimes, only one interpretation is possible because the phrase appears archaic, through a form that is obsolete or belongs to an older stage of the language. Most of the time, the archaism is lexical: some words that have now disappeared as free units still exist inside some idioms. These words are called fossils. For instance, *hue* is no longer used outside of the phrase *hue and cry*; it is the same for *fro* in *to and fro* (ancient *from*), or *shrift* (confession), which never appears without *short (to give somebody/something short shrift)*.

But before becoming opaque, those expressions were motivated. The creation of idioms is influenced by two main heritages: social heritage and cultural heritage. Even if the speaker perceives the world through his language, the latter is made up of things that exist in the speaker's social environment. Animals (*blind as a bat, as the crow flies, to take a hair of the dog that bit one, to kill two birds with one stone, to play possum*), insects (*a fly in the ointment, to have a bee in one's bonnet, to get ants in one's pants*), flowers and plants (*to gild the lily, to nip something in the bud*), vegetables and fruit (*to spill the beans, the apple of someone's eye*) are very present in idiomatic imagery. In the same way, if language mirrors life, men, and society, it also arises from culture. Several idioms originate in mythology (*in seventh heaven*—the antique firmament is composed of seven vaults, seven skies that one reaches successively, *to carry the weight of the world on one's shoulders, to rise from the ashes*), in the Bible (*not to know someone from Adam, to turn the other cheek, like a lamb to the slaughter, to cast pearls before swine, to wash one's hands off (something), to separate the wheat from the chaff*), in literature (*Achilles' heel, to bite the dust, the emperor's new clothes, it's all Greek to me*), and in history (*to have an axe to grind, when the balloon goes up, to haul someone over the coals, it's like the Black Hole of Calcutta, to win one's spurs, to cross the Rubicon*).

Because idioms cannot be understood literally, they cannot be translated literally either. The translation of idioms is characterized by a double problematics, as it is necessary to express not only the meaning and the form but also the aspect of fixation of the expression used in the source language. It is not always possible to meet this requirement. If the utterance in the target language is not a set phrase, the translation cannot be satisfying. When translating idioms, three different situations arise. First, the opaque expression can be translated literally because the same idiom exists in both the source language and the target language. This occurs with languages that are similar, such as English and French: *to read between the lines/lire entre les lignes*, *to praise to the skies/porter aux nues, a drop in the ocean/une goutte d'eau dans l'océan, to burn the candle at both ends/brûler la chandelle par les deux bouts*. Second, the idiom in the source language can have an equivalent idiom in the target language, like any other simple unit with its equivalent in a dictionary: *to have other fish to fry/avoir d'autres chats à fouetter, before you could say Jack Robinson/en deux temps trois mouvements, as alike as (two) peas in a pod/comme deux gouttes d'eau, the last straw (that breaks the camel's back)/la goutte d'eau qui fait déborder le vase.* Third, the idiom can be translated by a simple word or a paraphrase if neither the same idiom nor the equivalent exists in the target language. In this case, a part of the signified will be lost.

The expression of the same ideas varies from one language to another: when *there's something in the wind* in English, there's a conger eel under the rock in French (*il y a anguille sous roche*), and something is boiling in the pot in Italian (*qualcosa bolle in pentola*). If you are *in someone's good books* in English, you are in somebody's papers in French (*être dans les papiers de quelqu'un*) or in somebody's sleeve in Italian (*essere nella manica di qualcuno*). And if you *call a spade a spade* in English, you call a cat a cat in French (*appeler un chat un chat*), and you say bread to the bread and wine to the wine in Italian (*dir pane al pane e vino al vino*). Thus, idioms have a supplementary and particular aspect that simple words do not have: because they are motivated signs, and thanks to the images that are used, they give information about conceptions of the world considered by linguistic communities.

References

Bally, Charles. 1951. Traité de stylistique française. Geneva and Paris: Librairie Klincksieck.

Lakoff, George, and Johnson Mark. 1980. Metaphors we live by. Chicago: University of Chicago Press.

Nida Eugene Albert. 1964. Toward a science of translating. Leiden: E.J. Brill.

Vinay Jean-Paul, and Darbelnet Jean. 1963. Stylistique comparée du français et de l'anglais, (1995; Comparative stylistics of French and English: a methodology for translation). Paris: Librairie Marcel Didier.

CARINE MARRET

Igbo and Igboid Languages

Igbo (Ibo, in old orthographies) is one of the three major Nigerian languages. It is spoken by some 30 million people in present-day Imo, Abïa, Ebonyi, Enugu, Delta, Rivers, and Anambra states, and also in parts of Cross River, Akwa Ibom, Bayelsa, and Benue states of Nigeria. Igbo is now classified as belonging to the 'Benue-Congo', previously and variously to the 'Kwa', the 'Eastern Kwa'. The current Benue-Congo group includes not only Yoruba, Edo, and other smaller Nigerian languages but most of the 'Bantoid' languages of eastern, central, and southern Africa. The basis for this classification has been seriously questioned, and has been shown in many cases to be untenable. The proposition advanced by Kay Williamson (1989) that Igbo was a language of the lower Niger Delta from where it spread back north to present-day Igbo heartland has also been largely rejected by students of Igbo history and linguistics such as M. Angulu Önwüejiögwü (1980) or E. Nölue Emenanjö (2001).

Igbo is a single language, with several dialects. Following the Biafra/Nigerian war (1967–1970), there has been continuing pressure on some Igbo speakers to formally dissociate their dialects from Igbo proper. This has been particularly so with respect to Ikwere, Ahoada (Ekpeye), and Ïka. In fact, the term, 'Igboid', was coined in 1982 on an analogy with the other Benue-Congo languages (Edoid, Defoid, Platoid, etc.), in the face of an attempt to eliminate the 'Igbo' language name altogether and replace it with the nondescript term 'Lower Niger'. Still, the effect of the use of the term Igboid has been to make Igbo just one other member of a cluster of 'languages' previously understood to be dialects of Igbo—Agbo, Ekpeye, Ikwere, Izi, Ïka, Ogbah, and Ükwüanï/Enuanï. For these reasons, the term 'Igboid' is now disregarded as a strictly linguistic term since it misrepresents and understates the remarkable internal linguistic unity of Igbo among all the other languages of the Benue-Congo group.

Broadly speaking, allowing for the special dialects mentioned above, there are two major dialect zones in Igbo: Owere and Önïca, although the early missionaries proposed three: Isuama, Önïca, and 'Bonny' (Samuel Crowther 1882). In each of these zones, however, native speakers can identify sub-dialects, each marked by some distinctive detail of phonology, e.g. the glottal fricative in Ngwa; the voiced implosive /ɗ/ replacing the alveolar plosive, or the voiced glottal fricative /ɦ/ in Übakala, the ninth vowel in Izi. Other syntactical and lexical differences are also in evidence, of course; but as may be seen from the grammars of these dialect/languages (e.g. Meier and Bendor-Samuel 1975), the basic grammatical features remain the same for all Igbo dialects.

The history of Igbo language scholarship has been intimately tied up with the matter of its dialects. In its first efforts, the Church Missionary Society promoted 'Isuama Igbo' as the future national Igbo dialect, comparable to Oyo-Yoruba. The plan failed. The Roman Catholic Mission chose the Önïca dialect for its translations and other publications. This also proved unacceptable, as did 'Union Igbo', being promoted by the Colonial Education Department. A Standard Igbo (Igbo Izugbe) has now emerged, which combines generalized features of vocabulary and grammar from the Önïca and Owere dialects. This is the standard currently in use in the school system, and for most publications, although not without dissent from some eminent quarters.

The standardization of the Igbo orthography was key to this development. The first orthographies of the language were devised by repatriate Igbo missionaries of the Church Missionary Society, notably Rev. John C. Taylor, working under Bishop Samuel Crowther, who adopted Richard Lepsius's 'Standard alphabet' in a modified form for *Isuama primer* (1859, 1860), and the *Vocabulary of the Ibo language* (1882). A 'war of the orthographies', however, ensued when the Roman Catholic Mission began promoting its own system based largely on a pan-Nigeria alphabet developed by the International Institute of African Languages and Cultures (London) and recommended by the [Colonial] Education Board, Lagos in 1931. The resulting conflict between the Missions was not resolved until 1961, with the adoption of the Official (Önwü Committee) Orthography. This orthography has gained general acceptability, owing in large measure to the leadership of F.C. Ögbalü and the Society for the Promotion of Igbo Language and Culture, which also codified the basic elements of standard written Igbo. A New Standard Orthography introduced in 1998 adopts the Önwü Orthography, but replaces the subdotted vowels with their more readily available umlauted versions, as follows:

A, B, C/CH, D, E, F, G, GB, GH, GW, H, I, Ï, J, K, KP, KW, L, M, N, NW, NY, Ñ, O, Ö, P, R, S, SH, T, U, Ü, V, W, Y, Z.

Interestingly enough, although Igbo is a tone language using intonational pitch to differentiate meaning, most books (other than textbooks for teaching the language) are rarely tone-marked. The situation does not show any signs of changing any time soon.

What follows is a brief review of the phonological and syntactical features of Igbo based on Standard Igbo, with references to dialect variants as necessary.

Phonology

Igbo has eight phonemic VOWELS /a/, /i/, /ö/, /ü/; /e/, /i/, /o/, /u/. A ninth vowel (sometimes /a/, other times /ë/) is encountered in some dialects, notably Ïka, Ükwüanï, Önïca, and Izi. These vowels occur in two sets: Set (1): /e/, /i/, /o/, /u/, and Set (2) /a/, /i/, /ö/, /ü/ whose vowels, like the ninth vowel, are produced with an advanced tongue root. Igbo vowel harmony rules do not permit the co-occurrence of vowels from both sets within the same word.

The CONSONANTS include the stops (/b/, /p/, /d/, /t/, /j/, /ch/, /g/, /k/, /gw/ and /kw/); the implosives (/kp/, /gb/); the fricatives (/v/, /f/, /z/, /s/, and /sh/); the nasals (/m/, /n/, /ñ/, /ny/, and /nw/); the trill (/r/); the lateral (/l/); and the approximants (/y/, /gh/, /w/, and /h/). In some dialects, the bilabial plosives, the palatals, and labialized velars are often aspirated, the fricatives sometimes nasalized. Aspiration of the plosives has phonemic significance in certain dialects. Some fricatives and approximants are palatalized across dialects: /v/ > /f/ > /h/ > /hw/; or /h/ > /s/ > /sh/ > /r/, even /l/. Other variations include the /l/, /r/, and /n/ interchange in **ala/ana** (= earth); **ala/ara** (= breast), and the use of /nw/ for /w/, especially in Ngwa and Ikwere (e.g. **nwanyï/wanyï** = woman).

Tone System

Igbo is a two-tone (High, Low) language. The vowel bears the syllable tone. Tone is also a factor in the lexicon, morphology, and syntax. It is the only difference between **akwa** [LH] 'egg' and **akwa** [LL] 'bridge'; or between **saa** [HH] 'be wide or outspread' and **saa** [LH] 'answer; give reply to'. For nouns, the pattern of the final syllable determines their inherent tones. Most verb inflexions involve variations of tone. The primary distinction is still between verb forms with High or Low tones in their (first) root syllable. Since the infinitive prefix (**i/ï**) is always high, H-tone verbs may or may not accept low-tone inflected forms, whereas L-tone verbs remain low tone in all their inflected forms.

Downstep, a phonetic lowering of a high tone after another high tone, is a recurring feature of the language, and so also Downdrift in which there is a lowering of high tones after a high tone in a terrace-like movement over an entire word or utterance, irrespective of word boundary.

Syllabic Structure

The Igbo SYLLABLE consists of a vowel (V), a nasal consonant (/m/, /n/), or a vowel preceded by a consonant (CV). Unbound forms (other than pronouns) are either CV (as in **di** 'husband'; **ji** 'yam') or VCV (as in **ada** 'daughter'; **uru** 'grief'); or else, they are CVV (e.g. **gaa** 'go') where the final V is a marker for the imperative and is sometimes parsed as a verb stem + vowel suffix combination. Polysyllabic CVCV forms, except for initial nasal consonants (**mmanü** 'oil') and borrowings (**fada** 'priest') are invariably inflected verbs; polysyllabic VCV(V) forms are either simple nouns, or nominalizations from verbs form.

Syntax

The noun and the verb are the only significant grammatical class in Igbo, pronouns, adjectives, adverbs, and similar classes being either technically nonexistent, or having (in some instances) no more than two or three members. NOUNS and nominals (with a few exceptions) have an initial VCV syllable structure; further syllables may be added, in every case ending in a CV. General rules govern how two vowels may be adjacent to each other, and the conditions for assimilation and tone change between them. Nouns are not inflected. Polysyllabic nominals with initial nonnasal consonants (e.g. **gburugburu** 'all around; on all sides') are usually adjectivals, adverbials, or ideophones formed from base verbs. Nouns are never inflected for number or gender.

VERBS are the only inflected class of words in the language. The verb has a primary root, (usually derived from the infinitive form), and either one or more secondary roots from other bound verb forms, or from extensional suffixes. There are two verb types: (i) the regular verbs and (ii) the so-called auxiliaries (**dï, ga, ka, ma, na**) and copulas (**bü, dï, nö**, and **wü**). In compound (sometimes called complex) verbs (e.g. **dapuo** [<**daa** + **puo**] 'fall' + 'have a gap or opening', the end-verb carries the appropriate inflectional ending in affirmative imperatives.

TENSES include Present, Past, Future, or Past Perfect. The Present is actually an absolute present and only a few verbs can express it. The present tense form is the same as the verb root, without inflexional affix. The Simple Past is formed with an rV inflexional affix after the root, V being a reduplicate or copy of the verb root vowel. The Past Perfect is formed with an rV suffix attached not to the verb root, but to the affirmative imperative forms, almost always a CVV. The Future

tense is formed with an auxiliary **-ga** and a harmonizing **a/e** suffix after the verb root. The initial prefix is high, the second always low, except for verbs with HH tones in the imperative.

Aspect

ASPECT is distinguished from tense. In the present tense, aspectual inflexion may be used, for example, to reflect the duration of an action, the fact that it is ongoing, or that it is customary. Similarly, in the past tense, the harmonizing aspectual prefix **a/e** + verb root + inflexional suffix [**la/le**; **ne/na/**; **go/we**, depending on dialect] may be used to indicate fully completed (perfected) action.

Affixes, Suffixes, and Enclitics

INFLECTIONAL AFFIXES are always attached to the verb, either pre- or post-positionally, to specify tense or aspect with a harmonizing prefix **a/e**; a suffix **-we** (or dialect variants). A harmonizing rV suffix with the so-called stative verbs indicates continuing present time or the past tense. To confirm a negative status for a statement, the verb uses a harmonizing **ghï/beghï** and their dialect variants. EXTENSIONAL SUFFIXES (sometimes called lexical suffixes) are a major element in the syntax of the Igbo verb. (En)clitics are also usually affixed to verbs, but they are excluded from suffixes because they can also occur independently, as free morphemes in association sometimes with a noun phrase (NP), and other times with a verb phrase (VP).

Word Order

The basic WORD ORDER (in declarative sentences) is S(ubject)–V(erb)–O(bject)-[1]-O(bject)[2], or more exactly, N(oun)P(hrase)[1] + (VP [+ NP[2]]). The O[1] and O[2] nouns may be the direct object of the verb, or an inherent or cognate complement. Transitivity in the verb is difficult to establish because of the language's extensive use of verb complements—inherent, cognate, or otherwise.

Noun Phrases

The NOUN PHRASE, whether NP[1] or NP[2], is made up of the nominal head and the qualifiers that come after it, with the exception of the numeral, **otu** 'one', which always precedes the head. The modifier may be a word, phrase, or clause, in that order. Qualifiers come after the Igbo noun head.

Verbal Phrases

The VERB PHRASE (VP) is a complex inflected unit, marked by an initial (nonmandatory) auxiliary or a verbal prefix. This is followed by the verb stem and the required inflections (mood, tense, negation/affirmation) together with enclitics that extend the meanings of these verbs.

Situation Today

Igbo is the 'mother tongue' of some of Africa's eminent writers and artists: Chinua Achebe, Cyprian Ekwensi, Christopher Okigbo, for example. Still, original works in Igbo by these established writers are hard to come by. A kind of pan-Igbo cultural revival appears to be growing, however, and with it, a renewed effort to speak, write, and teach the language at all levels. Already several fictional works have been published for the school market and the general public. Two major dictionaries have been recently published (Echeruo 1998, Igwe 1999); others are in preparation. In addition, a major research effort is under way for the production of an archival Igbo dialect dictionary, in addition to the effort to create an Igbo metalanguage.

References

Achebe, Chinua. 2001. Echi dï Ime; Taa bü Gboo. Odenigbo Lecture Series, no. 3. Owerri: Assumpta Press.

Bennett, Patrick R., and Jan P. Sterk. 1977. South-Central Niger-Congo: a reclassification. Studies in African Linguistics 8. 241–73.

Crowther, Samuel. 1859;1860. Isuama-Ibo Primer. Revised and enlarged by Rev. John Christopher Taylor. London: Church Missionary Society.

Crowther, Samuel 1882. Vocabulary of the Ibo language. London: Society for Promoting Christian Knowledge.

Echeruo, Michael J.C. 1998. Igbo–English dictionary: a comprehensive dictionary of the Igbo language, with an English–Igbo index. New Haven: Yale University Press.

Emenanjö, E. Nolue. 1987. Elements of modern Igbo grammar: a descriptive approach. Ibadan: University Press Ltd.

Hair, P.E. 1967. The early study of Nigerian languages: essays and bibliographies. Cambridge: Cambridge University Press.

Igwe, G. Egemba. 1999. Igbo–English dictionary. Ibadan: University Press, Plc.

Meier, Paul, and Inge, and John Bendor-Samuel. 1975. A grammar of Izi: an Igbo language. Norman, OK: Summer Institute of Linguistics, University of Oklahoma.

Nwachukwu, P.A. 1987. The argument structure of Igbo verbs. Lexicon Project Working Paper, 18. Cambridge, MA: MIT Center for Cognitive Science.

Nwachukwu, P.A. 1981. Towards an Igbo literary standard. London: Kegan Paul.

Oluikpe, Benson. 1978. Igbo transformational syntax: the Ngwa example. Onitsha: Africana-FEP Publishers Ltd.

Önwüejiögwü, M. Angulu. 1977. Some fundamental problems in the application of lexostatistics in the study of African languages. Oduma 3(2). 29–36.

Önwüejiögwü, M. Angulu. 1980. An Igbo civilization: Nri kingdom and Hegemony. London: Ethnografica.

Williamson, Kay. 1989. Benue-Congo Overview. ed. by John Bendor-Samuel, The Niger-Congo languages, 247–74. Lanham, MD: University Press of America.

MICHAEL J.C. ECHERUO

India

India represents an astonishing array of linguistic diversity with four language families, and more than 600 languages as well as numerous geographical, social, ethnic, religious, and rural varieties or dialects. The subcontinent competes with Papua New Guinea in terms of the sheer numeric weight of its languages and varieties, a rural–urban dichotomy, and varied social structures. For these reasons, India is often described as 'the linguistic laboratory of the World'. Consequently, one finds a complex array of linguistic situations in terms of the extent of bilingualism, ranging from diglossic bilingualism to bilingualism based on three of the five highest ranking languages in the World (Hindi, English, and Bengali) according to the number of speakers. The magnitude and scale of linguistic diversity and parameters of language usage are often beyond the comprehension of speakers who are accustomed to western-style monolingualism.

Linguistic Profile

Approximately one in every six inhabitants of the world (that is 1.0 billion people) speaks at least one of the Indian languages. Indian languages belong to four language families: Indo-European, Dravidian, Austroasiatic, and Tibeto-Burman. Indo-European, with 12 major and three minor languages, represents the most influential language family both in terms of the number of speakers and social significance.

The 12 Indo-Aryan languages of the Indo-European language family are: Hindi, Assamese, Oriya, Marathi, Konkani, Gujarati, Urdu, Punjabi, Sindhi, Bengali, Kashmiri, and Nepali. Hindi and Bengali are ranked second and fifth in the world, respectively, in terms of the number of speakers, according to the World Almanac (2002: 447). Approximately 80% of the population of India speak languages of the Indo-European stock. Kashmiri (India and Pakistan) belongs to the Dardic sub-branch of Indo-Aryan. The three minor languages are as follows: Shina, Dogri, and Bhili. The Indo-European languages not only dominate India but also all the countries of the subcontinent (i.e. Pakistan, Nepal, Sri Lanka, and Bangladesh).

The second major family are the Dravidian languages, which include four major languages: Tamil, Telugu, Malayalam, and Kannada. These four languages are spoken in South India. Tamil is spoken both in India and Sri Lanka.

Two branches of the Austroasiatic language family are also found in India: the major members of the Munda subfamily are Santhali, Mundari, and Ho, which are spoken in southern Bihar, northeastern Andhra Pradesh, and broader areas of West Bengal in India. The Mon-Khmer branch is represented by Khasi, which is spoken in Meghalaya state of India.

Tibeto-Burman is represented by many languages, spoken in the subcontinent's northeastern borderlands. While Tibetan or Bhotiya languages are spoken in India (in Laddakh and Sikkim), Bhutan, and Nepal, the major concentration of the Gurung-Tamang group is present in central to eastern Nepal.

Nineteen languages are recognized by the Indian constitution. In addition, there are numerous dialects and languages. The number of vernaculars/dialects reported ranges from 216 to over 1,500 (1,652 mother tongues self-reported in the 1961 census; 216 mother tongues with more than 10,000 speakers each reported in the 1991 census).

The linguistic situation and communication networking in India can be represented diagrammatically by a pyramid-type structure, shown in Figure 1.

Languages and Modes of Communication

Hindi and English are the two national and link languages of India. While Hindi is the language of the masses in the northwestern and north-central part of India, English is the pan-Indian language of the educated elite. Thus, English and Hindi represent the peak of the pyramid.

As noted earlier, Hindi, even without combining with Urdu, has the second largest number of speakers among the languages in the world and is the lingua franca of India and of the subcontinent. The great Moghul empire in the past, and Hindi films and the mass media in the twentieth century, have played a significant role in the spread of Hindi-Urdu, not only in India but also across the subcontinent. Hindi and Urdu are mutually intelligible but are written in two different scripts—Hindi is written in the Devanagari script and Urdu is written in Perso-Arabic script. There is also a difference in literary affiliation—Urdu is associated with Persian and Arabic literary traditions, while Hindi is tied to the Sanskrit tradition indigenous to India. In addition, Urdu borrows its technical and learned vocabulary from Persian and Arabic, whereas Hindi borrows the same kinds of terms from Sanskrit.
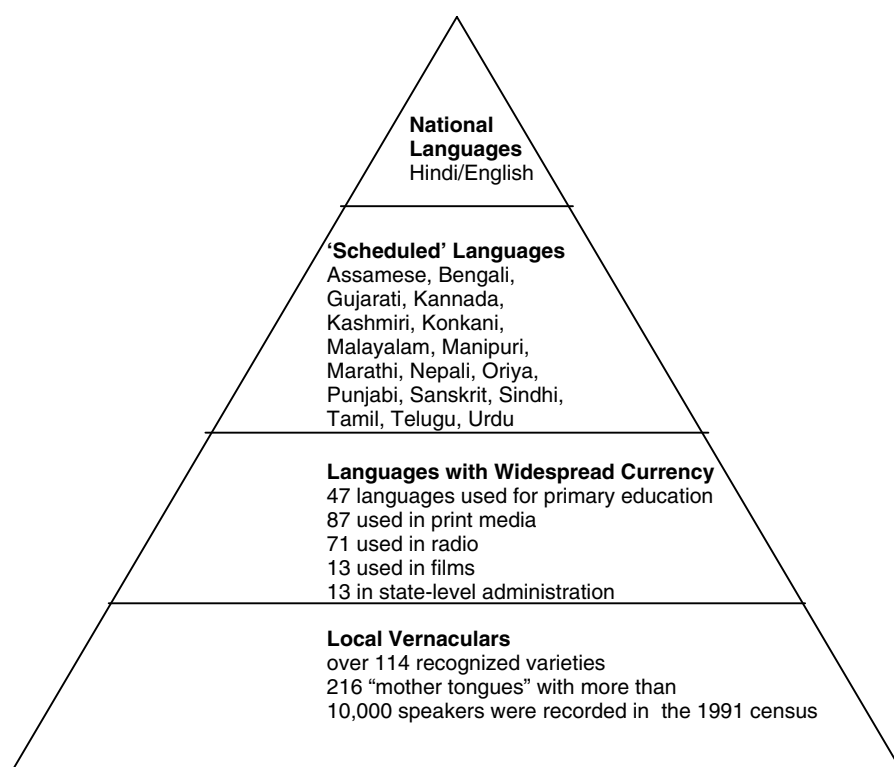
Figure 1. Linguistic situation in India.

'Scheduled' languages (henceforth, state languages) are spoken predominantly in their respective states: see Table 1 below. Hindi, along with English, is the only state language that is spoken in more than one state. Urdu is the official language of the state of Jammu & Kashmir. However, not all 'scheduled' languages are spoken in a particular state. Sindhi is not the official language of any Indian state. Kashmiri is spoken in the state of Jammu & Kashmir but the official language of the state is Urdu. Finally, rural dialects (mother tongues or vernaculars) are shown at the base of the pyramid.

As pointed out earlier, Modern Tamil, Telugu, Kannada, and Malayalam are Dravidian languages and are primarily spoken in the South, while Indo-Aryan languages, the descendants of the oldest documented Indo-Aryan language, Sanskrit, are primarily spoken in the North. Dravidian languages are the descendants of the proto-Dravidian languages.

Sanskrit is also recognized as an official language of India and is the language of India's cultural and intellectual tradition. English is the language of modern intellectual communication and is the official language of three states and all 'union territories'.

Before continuing with a discussion of the language situation in India, it is worthwhile discussing the sources of figures in Table 1—the census of 1991. Because the census asks for self-report of languages, it

should be noted that there are a number of problems with these figures (and other census figures as well). A claimed language may be a language or a nonlanguage—a reflection of regional, religious, caste, social, ethnic, literary and script affiliation, and even occupation. For example, it is not unusual for an ironsmith to report Lohari 'ironsmith' as his mother tongue or for a villager to report a dialect as simply the name of his village or caste. Alternatively, a language might be a speaker's perception of language/dialect on a language continuum, thus blurring language vs. dialect; or language vs. language distinction (in language contact situation).

The label 'mother tongue', which is the term used by the census, may also be ambiguous. An individual questioned may interpret this term as referring to his or her cradle language, parent tongue, language ordinarily used, or prestige language only. A prestige language such as English is often over reported, while stigmatized varieties (e.g. immigrant languages such as Saraiki from Pakistan) are underreported. An illustration of these problems is the fact that almost 50,000 speakers reported Sanskrit as their mother tongue (see Table 1). It seems extremely unlikely that Sanskrit is the cradle language of anyone. However, its prestige and daily use by some Indians have apparently led them to claim it to be their mother tongue.

TABLE 1   Scheduled Languages of India

Name	Language Family	State/spoken in	Number of Mother Tongue Speakers (1991 Census: Govt. of India)
Assamese	Indo-Aryan	Assam	13,079,696
Bengali	Indo-Aryan	Bengal	69,595,738
English[1]	Indo-European; Germanic	Meghalaya, Nagaland, Tripura	178,598
Gujarati	Indo-Aryan	Gujarat	40,673,814
Hindi	Indo-Aryan	Hindi belt: Bihar, Chattisgarh, Rajasthan, Haryana, Delhi, Himachal Pradesh, Jarkhand, Madhya Pradesh, Uttar Pradesh,	337,272,114
Kannada	Dravidian	Karnataka	32,753,676
Kashmiri	Indo-Aryan	Jammu & Kashmir[2]	56,690
Konkani	Indo-Aryan	Goa	1,760.607
Malayalam	Dravidian	Kerala	30,377,176
Manipuri	Tibeto-Burman	Manipur	1,270,216
Marathi	Indo-Aryan	Maharashtra	62,481,681
Oriya	Indo-Aryan	Orissa	28,061,313
Punjabi	Indo-Aryan	Punjab	23,378,744
Sanskrit	Indo-Aryan	No state	49,736
Sindhi	Indo-Aryan	Metro areas of western India	2,122,848
Tamil	Dravidian	Tamil	53,006,368
Telugu	Dravidian	Andhra Pradesh	66,017,615
Urdu	Indo-Aryan	Jammu and Kashmir	43,406,932

[1]The number of mother tongue speakers of English decreased in 1991! According to 1981 census, the mother tongue speakers of English were 202,000.

[2]These figures are not complete as no census was taken in J&K in 1991. Urdu and other scheduled languages are also spoken outside their respective states.

Sources and Processes of Multilingualism: Role of Sanskrit, Persian, and English

Three millennia of language contact have led to the convergence of four language families in the form of Aryanization of Dravidian languages and Dravidianization of Aryan languages. In addition, Persian, which was the official language of India during the Moghul empire, and the newest arrival, English, have added greatly to the mixed character of Indian languages, particularly Hindi. Hence, it is not uncommon for a Hindi speaker to switch between Persianized Hindi, Sanskritized Hindi, and Anglicized Hindi. In short, three contact languages have played a particularly important role in the history of India: Sanskrit, Persian, and English. Each has served as a prime vehicle for language contact and bilingualism, both historically and in the present day.

Sanskrit

It is well known that Sanskrit language is the single major force responsible for the formation of Indian languages and literatures. So great is the influence of Sanskrit on Indic languages that almost every literary and language-related activity is affected by it. As the language of the Hindu scriptures, and of the epic and classical literature, Sanskrit has traditionally constituted the medium of philosophical and technical literature throughout South Asia. A number of other factors such as past royal patronage and the promotion of Sanskrit education by the Brahmins from the sixth century BC onward have reinforced its prestige. Consequently, Sanskrit has played a leading role not only in the process of Aryanization of Dravidian and other languages but also in forming the ethos of India and other South and South East Asian countries to a varying degree.

For example, the effects of borrowings from Sanskrit on the lexical and grammatical systems of Dravidian languages have been profound. Murray Emeneau and Thomas Borrow (1962) note the tendency 'for all four of the Dravidian literary languages in the South to make literary use of the total Sanskrit lexicon indiscriminately'. So massive has the influence been that it is hard to utter more than a few sentences in these languages without using a word borrowed from Sanskrit. Over a century ago, Western scholars subscribed to the view that the Dravidian languages descended from Sanskrit, although Robert Caldwell, in his classical work, dispelled this misconception. About 20% of the noncultural part of basic vocabulary in literary Dravidian languages is loaned from Indo-Aryan mostly from Sanskrit. This proportion peaks to 50–60% in some languages due to shared cultural beliefs (e.g. Brahmanic values). This resulted not only in the replacement of content and function words but also had a substantive impact on the phonological, morphological, and syntactic structures of the Dravidian languages.

Furthermore, the study of *VyaakaraNa* (grammar) in Sanskrit was viewed as an area of the highest linguistic sophistication and a testing ground for various linguistic theories. The Sanskrit grammatical tradition resulted in grammatical works such as Panini's grammar, the *Ashtaadhyaayii,* which is still regarded as 'one of the greatest monuments of human intelligence' not only in the history of Indic grammatical tradition but also in linguisitics.

The Dravidianization of Indo-Aryan languages also took place, although the impact of Dravidian languages on Indo-Aryan languages was not nearly as extensive; nevertheless, this mutual influence (on lexicon as well as grammar) set the stage for leaky boundaries among the languages of India whether they were genetically related or not.

Not only this, Sanskrit also become a marker of caste identity, which in turn gave rise to diglossic bilingualism through Brahmin speech in Dravidian languages, particularly in Tamil. The use of Sanskrit in topical domains such as religion, philosophy, poetics, science, technology, and mathematics in Dravidian, as well as in Indo-Aryan-speaking areas, is particularly noteworthy. Hence, Sanskrit became the single most important marker of Indian culture—both in the north and in the south.

The role of Sanskrit in promoting bilingualism and language modernization inside and outside of India (e.g. in South East Asia) to this day is particularly striking. Today, the All India Radio relays news in Sanskrit and even newspapers are published in the language. Just as terms for technological innovations are created from Latin and Greek for western languages (*television*= Greek *tele-*'far', Latin *–vision* 'seeing'), Sanskrit is the source of designative cultural and technical innovations in South Asia and beyond. Consider, for example, that the technical vocabulary is drawn either from Sanskrit or English: the terms for prime minister (*pradhaan mantrii = pradhaan* 'head, principal'+*mantrii* 'minister'), president (*raashTrapati = raashTra* 'country' + *pati* 'husband'), radio (*aakaashvaaNii = aakaash* 'sky' + *vaaNii* 'voice, speech'), and TV (*duurdarshan + duur*(Hindi) 'far' + *darshan* 'sight'). The words given in the parenthesis are drawn from Sanskrit. Furthermore, Sanskrit plays an important role in the process of language modernization of South East Asian languages.

Persian

Language contact among Persian, Indo-Aryan, and Dravidian languages (Kannada and Telugu) formally began with the annexation of Punjab by the Turkish ruler Mahmud Gaznavi (twelfth century). The language contact situation with Persian was different from the Sanskritization of Indian languages. Persian was viewed as an imposed language in much of India; thus, with Moghul patronage, Persian became a marker of Islam. As a consequence, after independence in 1947 when India and Pakistan became independent nations, the influence of Persian continued to regenerate in Pakistan, a Muslim nation, while it reached a state of fossilization in India. In contrast, after independence, Sanskrit began to regenerate its original impact in India due to its association with Indian culture (particularly with Hindu religion). Just as extensive Sanskritization played an important role in the separation of Malayalam from Old Tamil, extensive Persianization together with Arabicization led to the separation of Urdu in Pakistan from Hindi-Urdu in India. Similarly, heavy Persianization and Arabicization of Hindi led Western grammarians like George Hadley (1772) and others (Schultz 1745) to claim that Persian must be the mother tongue of Hindi. Others thought that the ancestor of Hindi must be a Semitic language, perhaps Arabic or Hebrew. The concept of Indo-Aryan languages and the genetic hypothesis of languages was still a distant reality in the nineteenth century. As pointed out earlier, Hindi and Urdu in their spoken style are essentially the same language, approximately parallel to British and American English.

Persianization of Indian languages was not restricted to the lexicon (content and function words). The impact of Persian borrowing on the grammar of Indian languages was as significant as the influence of Sanskrit. The morphological processes such as reverse *compounding (sher-e-panjaab* 'the tiger of Punjab' rather than the unmarked pattern—*panjaab kaa sher* 'Punjab's tiger), inflectional morphology (plural markers), and word compounding with Persian

became part of modern Indo-Aryan languages and so did conjunct verb construction, complementation (with *ki* 'that'), and conjunction with noun phrases.

Domains such as law and regulation, sports, and business belong to Persian-mixed Hindi to this day in India. A highly Persian-Arabic style is widely used in the modern-day legal register. Also, newspaper reporting, a business section that includes a Price-index report, share market, and reports dealing with economics such as budgetary reporting, and so on, incline toward the Persianized style.

English

Although English is a newcomer in the linguistic scenes of India, it has already registered some significant results from the viewpoint of the sociology of communication, in general, and cross-linguistic communication, in particular. At present, it serves as one of the most important sources of pan-Indian bilingualism or pan-Indian communication. Consequently, it has added greatly to the already mixed character of Indian languages. What is noteworthy is that mixing with English is an important linguistic feature of India. When the introduction of English to the Indian linguistic landscape began with the dawn of the British colonial era, English began to develop roots in Indian education. A blueprint for India's educational policy was laid down in Lord Macaulay's *Minute* (Feb. 2, 1835). Macaulay's stated mission for the British Raj of creating 'a class of persons, Indian in blood and color, but English in taste, in opinions, in morals and intellect' introduced English education to India; his attitude not only toward Indian languages but also toward the languages of Southwest Asia—particularly, Arabic.

> I have no knowledge of either Sanskrit or Arabic. But I have done what I could to form a correct estimate of their value...I am quite ready to take the oriental learning at the valuation of the orientalists themselves. I have never found one amongst them who could deny that a single shelf of a good European library was worth the whole native literature of India and Arabia.

More than one and a half centuries later, English has overcome its status as merely the language of the colonial power, and has become an integral part of the Indian linguistic mosaic. Contrary to the most popular preindependence consensus that Hindi would dethrone the English language after independence, English not only continued to flourish in the educational and official network of India but also became one of the official languages of the nation and thus continues to enjoy the patronage of the Indian elite. Although the incidence of bilingualism with English is still very low, English has acquired domains such as education, law, government, media, and science and technology, which

once belonged to either Sanskrit or Persian. The uses of English, parallel to its predecessor contact languages— Sanskrit and Persian—has led to the Englishization of Indian languages. On the other hand, English has undergone significant changes locally to carry the communicative burden of Indian society. The process of nativization of Indian English continues to this day.

English has colored the phonology, morphology, and syntax of Indian languages. Hindi, for instance, has borrowed several consonant clusters such as *st-, sk-, sl-*, and phonemic segments such as /f/ and /z/ as the result of large-scale infiltration of English lexical items, such as station, school, slow , fail, etc. Even at the level of language, which is less prone to change, i.e. syntax, Indian languages have not managed to escape the influence of English. The preference for passive over active construction is attributed to the influence of English.

Turning to the first layer of the pyramid, along with English, Hindi provides a major contact language link within the communicative network. The heart of this link consists in the Hindi-Hindustani-Urdu-Punjabi core/axis. This axis forms a giant speech community with direct links to Bengali in the East, Gujarati and Marathi in the West, and Telugu and Kannada in the South. In contact with these languages and other regional varieties, Hindi has developed its own regional varieties, for example: Mumbai Hindi (Bombay Hindi, Bindi), Kalkatiya Hindi (Calcutta Hindi), Madrasi Hindi (Madras, renamed Chennai, Dakkani Hindi 'southern Hindi').

Other factors such as education, military, migration, literature (particularly devotional literature), religion (pilgrimages), popular and electronic media, and trade play an important role in societal bilingualism or multingualism in India.

Education and Government Policies

The language policies of the government of India are very conducive to the promotion of the language rights of minority languages and the advancement of linguistic diversity and pluralism. This is evident from the facts regarding languages represented in the third tier of the pyramid. The number of languages used in primary education is 47; 87 in the print media; 71 in radio and broadcasting; 13 in films; and 13 in state-level administation. In addition, the introduction of the Three Language Formula in education is yet another notable feature of national policy. This formula calls for trilingualism or quadrilingualism in education. In addition to learning of the two national languages— Hindi and English—students are expected to learn a third language apart from their native tongue. For example, in the Hindi-Urdu-Punjabi belt, students are expected to learn one of the four Dravidian languages

(Tamil, Telugu, Kannada, and Malayalam). Although the debate concerning the effectiveness of the Three Language Formula is still brewing, the underlying merit of the educational policy in the promotion of mulitilingualism is hardly questionable and best represents the multilingual character of the nation. Similarly, the National Academy of Letters (The Sahitya Akademi) of the Government of India and its regional counterparts promote literary activities at least in the 19 'scheduled' languages.

Bridging the Rural vs. Urban Divide: Literature, Popular Media, and Trade

Unchanged for centuries, rural populations constitute the heart of India. According to the 1991 census, for instance, most of the population of India (about 77–78% of the total population) lives in more than half a million (627,000) villages and speaks in numerous vernaculars (see the base of the pyramid in Figure 1). The most obvious linguistic vehicles for reaching rural India are either Hindi or the regional languages and their local vernaculars. The incidence of literacy in English is not significant in rural India. Thanks to literature in the past and popular media at present, the boundaries between rural vernaculars and Hindi have become very fluid. Historically, literature (e.g. the devotional poetry tradition—poets such as Kabir, Tulasi, Surdas, and Meera Bai) has played a major role in neutralizing such boundaries and bringing the influence of regional languages (from East to West and South to North; bidirectional) and rural varieties into Hindi. The consequence was a mixed speech, which is termed 'sadhukari bhaashaa' (the language of saadhus and saints free from any prescriptive norms). At present, the most powerful and vital force for bridging the gap between the urban vs. rural (and regional) divide is the Hindi film industry and the mass media. These sources of influence offer a unique appeal in terms of cinematic techniques, dance, drama, and music, and even constitute a viable marketing alternative to Hollywood in the world of entertainment in India and outside India. The reach of the Hindi media indeed extends well beyond one sixth of the world's population that inhabits India, reaching members of the Indian diaspora worldwide. Consequently, mutual intelligibility between regional languages and rural varieties of Hindi is growing steadily. At least this is true of spoken Hindi—caltii Hindi 'colloquial Hindi' or bazaar or Dakkini Hindi. Nevertheless, some barriers do remain.

Salient Features of Indian Multilingualism

As is self-evident from the above discussion, it is clear that linguistic diversity is a hallmark of India. India is often labeled the 'Tower of Babel', since it is remarked that India 'babbles' chaotically in 'hundreds of dialects'. Such labels or expressions reflect a lack of understanding of the complex but structured network of linguistic communication and bilingualism in India. The main focus of this section is to highlight the results of prolonged and recent bilingualism in the region.

Indian Multilingualism: A Natural Phenomenon

There is no doubt that there are linguistic rivalries and conflicts in India, even language vs. dialect conflict. Centuries-old coexistence and an ongoing process of convergence have led to an unmarked pattern of widespread naturalistic linguistic convergence, rather than separation, dominance, and disintegration. Large-scale diffusion of linguistic features across genetic and areal boundaries has resulted in mutually feeding relationships and reciprocity. Although the incidence of 'antagonistic bilingualism' of the sort witnessed in Belgium and other parts of the world cannot be ruled out, its incidence is very limited in degree and scope. In short, one of the defining features of India is one of long-term stable bilingualism and linguistic accommodation. This situation led Caldwell (1903) to propose implicitly that India as a 'linguistic area'—that is, an area in which genetically distinct languages show a remarkable level of similarity and diffusion at the level of grammar. For instance, retroflex sounds (produced with the tip of the tongue curled upward) are the integral part of the languages of India and are thus dative subject constructions:

> mujhe kitaab pasand hai
> me   book liking is.
> 'I like the book'.

Language Attitudes and Linguistic Accommodation

Why has diffusion between genetically distinct languages taken place in India to a degree rarely witnessed in other parts of the world? The answer lies not only in language-internal developments but also in linguistic attitudes that are in turn associated with the 'worldview' of Indians since ancient times. William Bright (1984) characterizes linguistic attitudes in India as an 'accepting attitude, which has brought about the assimilation of features from Dravidian, Indo-Aryan, Islamic, and even Christian and European cultures into a single system, complex, but integrated'. Thus, linguistic accommodation is another important feature of Indian bilingualism.

Stable multilingualism and multiple identities

Multilingualism is not a borderline phenomenon in India nor is it restricted to either educated or business communities. Multiple languages and multiple language identities are defining features of Indian

multilingualism that reveal the dynamics of language usage and a constant negotiation of identities. For instance, the simple act of a social encounter and associated greetings is likely to begin with the choice of one of three modes of greetings, which reveal the religious affiliation of speaker, listener, or both in addition to conveying the social ritual of greetings: *namaste* or *namaskaar* (Hindu), *aadaab* (Muslim), *sat sri akaal* (Sikh), drawn from Sanskrit, Perso-Arabic and Punjabi, respectively. Not only this, a Hindu speaker may decide to use *raam-raam* and *jai maataa dii* to express regional (rural) and ethnic affiliation, respectively, within Hinduism. Similar conditions hold true of Muslims and members of other religious and social groups.

Verbal Repertoire

Code-mixing and code-switching are natural phenomena in the life of a bilingual and are thus all-pervasive in India. Bilingual societies organize their linguistic repertoire differently from monolingual societies. No one language is viewed as suitable for all communicative occasions. This has led to the emergence of mixed languages such as Hinglish (Hindi+English). The extent of the impact of English is so great that it has caused a great deal of alarm among purist speakers of Indian languages and is seen as threatening their independent existence.

The phenomena of code alternation have brought profound changes to the formal grammar of Indian languages. Consider, for example, the case of Causatives in Hindi. Hindi Causatives are derived by the morphological process of suffixation. By adding the suffixes *-aa-* and *-waa-*, the first and the second causative verbs are formed as in: *paR* 'to study/read', *paRaa* 'to teach' (lit. 'cause to study'), and *paRwaa* 'cause x to cause y to study/read'. Mixing with English has resulted in the introduction of a new class of transitive/causative verbs of a mixed type, i.e. *study karnaa* 'to research (on a topic); lit. *study+ karnaa* 'to do study'), *teach karnaa* 'to teach', *study karwaanaa* 'to guide research', *teach karwaanaa* 'to guide teaching'. The verb *study karnaa* 'to research' is not just a paraphrase or translational equivalent of Hindi *paR* 'to study/read'; the English-based verb expresses the meaning 'to research or study a topic from the viewpoint of research' and the causative counterparts, i.e. *study karwaanaa* 'to guide research', *teach karwaanaa* 'to guide teaching' highlight the 'facilitative' (i.e. help the causee to do a particular act) while the Hindi causative marks 'compulsive' meanings.

Before the introduction of English, Hindi had two conjunct verb expressions with the meaning 'to travel': *yatraa karnaa* and *safar karnaa*. The former is Sanskrit-based and the latter is Persian-based. The former con-notes leisurely travel with religious overtones, whereas the latter marks any ordinary travel. The English-based *tour karnaa* has added yet a new semantic dimension, which expresses the concept of business travel which is contrary to what the verb *tour* conveys in English.

Freedom of choice

Linguistic diversity highlights freedom of choice of speech, which is not just a recent, twentieth-century phenomenon in India, but a centuries-old tradition. This attitude has not been seen in many parts of the world. For example, French, Chinese, and some other major languages of the world were imposed on others as part of a movement to eradicate rural and regional language varieties.

Scripts

Indic scripts constitute yet another example of 'India as a linguistic area'. Both Indo-Aryan and Dravidian languages are written in various scripts, which descend from the Brahmi script (third century AD). Indian scripts are phonetic in nature and there is a fairly regular correspondence between the letters and their pronunciation. A number of Indian languages such as Hindi, Marathi, Nepali, and Sanskrit are written in the Devanagari script, whereas it is not uncommon to write a language in different scripts. For example, Punjabi is written in at least three scripts: Gurmukhi, Devanagari, and Perso-Arabic.

References

Annamalai, Elayaperumal. 2001. Managing multilingualism in India: political and linguistic manifestations. New Delhi: Sage Publications.

Bhasham, Arthur L. 1954. The wonder that was India. New York: Grove Press Inc.

Bhatia, Tej K. 1982. English and vernaculars of India: contact and change. Applied Linguistics 3(3). 235–45.

———. 1985. Religious-colonial models of language and early Hindi grammars. Lingua 65. 123–34.

———. 1999. Advertising in rural India: language, marketing communication and consumerism. Tokyo, Tokyo Press.

Bright, William. 1960. A case study of caste dialect in Mysore. Indian Linguistics 21. 45–50.

Bright, William, A.K. Ramanujan. 1964. Sociolinguistic and language change. Proceedings of the Ninth International Congress of Linguistics.

Caldwell, Robert P.A. 1903. A comparative grammar of the Dravidian or South-Indian family of languages. Madras: Oriental Books.

Emeneau, Murray B., and Thomas Borrow. 1962. Dravidian borrowings from Indo-Aryan. University of California Papers in Linguistics, 26. Berkeley: University of California.

Gumperz, John. J, and R. Wilson. 1971. Convergence and creolization: a case from the Indo-Aryan/Dravidian border in India. ed. by D.H. Hymes, Pidginiztion and creolization of

languages, 151–67. Cambridge: Cambridge University Press.

Hadley, George. 1772. Grammatical remarks on the practical and vulgar dialect of Indostan language. Menston: The Scholar Press.

Kachru, Braj B. 1983. The Indianization of English: the English language in India. Delhi: Oxford University Press.

Kachru, Yamuna, and Tej. K. Bhatia. 1978. The emerging 'dialect' conflict in Hindi: a case of glottopolitics. International Journal of the Sociology of Language 16. 47–56.

Khubchandani, Lachman M. 1997. Revisualizing boundaries: a pluralism ethos. New Delhi: Sage.

Krishnamurti, Bh., and Aditi Mukherjee (eds.) 1984. Modernization of Indian languages in news media. Hyderabad, India: Department of Linguistics, Osmania University.

Pandit, Prabodh B. 1972. India as a sociolinguistic area. Poona: University of Poona.

Schiffman, Harold, 1999. South and Southeast Asia. ed. by J.A. Fishman, Language and ethnic identity 431–43. Oxford, Oxford University Press.

Schultz, B. 1745. A grammar of the Hindostan language. Saxony: Hall.

Sridhar, S.N. 1981. Linguistic convergence: Indo-Aryanization of Dravidian languages. Lingua 5. 199–220.

TEJ K. BHATIA

See also **Bilingualism; Dravidian; Farsi; Hindi-Urdu; Indo-European 3: Indo-Iranian Languages; Language: Contact—Overview; Panini; Sanskrit**

Indian Ocean Creoles

In the Indian Ocean, several French-lexicon creoles are spoken by between one and two million people. They are usually divided into two mutually unintellibible and probably (largely) historically unrelated groups: (1) Réunionnais (Réu) of Réunion island, and (2) the so-called Isle de France (IdF) creoles, which emerged in Mauritius (Mau), and later (eighteenth to nineteenth centuries) spread to the Seychelles (Sey), Rodrigues, the Chagos Islands, and elsewhere.

Réu differs both historically and structurally from other creoles in several respects. Early settlement of the island was characterized by a relatively strong component of French speakers, and a rather slow influx of alloglotts. Whites made up a majority of the Réunionnais population until about 1715, i.e. half a century after the beginning of permanent settlement of the island. As late as in 1690, only one family in six did not include a native Francophone. Moreover, in search of subsistence, many members of the white proletariat were forced to take up employment in the fields alongside (former) slaves. As opposed to other plantation societies, where the economic power was always in the hands of a small white élite, a large proportion of the white population on Réunion has always been as poor as their dark-skinned compatriots. Together with a high rate of inter-racial marriages, this led to a language less distant from French than most creoles. In fact, most non-Francophone linguists see Réu as a semicreole (comparable to, e.g. Afrikaans) rather than as a creole proper, while in the French-speaking scholarly tradition, it is treated as a fully fledged creole on a par with e.g. Haitian.

As opposed to IdF, Réu maintains significant parts of French verbal morphology. Réu verbs have up to five different forms, some even having suppletives. The language, furthermore, has an attributive copula inherited from French, lacks bimorphemic interrogatives, and has very little of the incorporation into nouns of etymological articles, so characteristic of French creoles in general. The postverbal placement of the sentence negation also aligns it closely with its lexifier. Moreover, Réu retains grammatical gender, as opposed to virtually all (other) creoles. Some varieties of Réu also retain front rounded vowels.

It should be noted, however, that claims regarding the structure of Réu are difficult to make, and when made, they should be treated with some caution, for the island is dialectally very fragmented, with remarkable variation in speech from one area to another. The fact that there is a continuum between regional French and Réu among the island's almost 750,000 inhabitants also makes it impossible to give the number of speakers of either.

In contrast to Réu, the IdF varieties are typologically rather similar to Caribbean creoles. There is some evidence that a stable IdF had emerged within 50 years after the French settlement of Mauritius in 1721.

Although later slave imports were predominantly from Madagascar, and later East Africa, West Africans (speaking e.g. Wolof, Bambara, and Fon) were among the first arrivals in Mauritius, who may have contributed to making the language so similar to Caribbean creoles. Other substrate languages include Bengali and Tamil.

Occupied by the British for 150 years, English is Mauritius's official language, but the position of French (L1 of a small group) remains remarkably

strong at all levels, and it may in practice be said to be co-official. Mau, despite its lack of any official recognition whatsoever, is nowadays the home language of three fourths of the total population of 1,200,000. It is gaining ground in the Indian and Chinese languages introduced by nineteenth century immigrants whose descendants now make up the majority of the population.

The Seychelles, now with a population of about 80,000 was colonized in the 1770s, with settlers coming both from Mauritius and Réunion. Ethnolinguistically, the Seychelles differs from Mauritius mainly in having no sizable minority of Indian descent, and so far as native language is concerned, it is virtually monolingually creolophone. The Republic of the Seychelles is unusual in giving official recognition to its Creole language—the country is officially trilingual in Creole, English, and French. English is spoken as an L2 by most, and French by a sizeable minority. Media are available in all three official languages.

The IdF varieties are all mutually comprehensible without much effort. The most noticable differences between Mau and Sey are the use of the predicate marker *i* in the latter, and nasal vowels in Sey often corresponding to oral ones in Mau. In both cases, Sey is conservative vis-à-vis Mau. Lexically, Sey displays some influence from Réu in, e.g. the use of *komela* 'now' instead of *aster* as in Mau.

As mentioned above, IdF is rather similar to the better-known French creoles of the Caribbean, which may or may not have something to do with the early presence of West Africans in Mauritius. The main phonological difference is that postalveolar fricatives are depalatalized in the Indian Ocean (i.e. /ʃ, ʒ/→/s, z/). IdF also has only five oral vowels, as opposed to seven in the Caribbean creoles.

The system of preverbal tense–mood–aspect markers is particularly close to that of, e.g. Haitian, with progressive *(a)pe* (< *après*), past *ti* (< *te* < *était/été*), completive *((f)i)n* (< *finir*), and two different futures *(v)a* (< *va*) and *pu* (<*pour*).

Pluralizing (see below) and overt copular (ete < *été*) morphemes differ from those used in the Caribbean, but both appear to be postformative developments. Another such example is the verb apocope (*mãze* vs. *mãz* 'to eat'), which occurs wherever the verb governs an overt element.

Also noteworthy is the exceptionally high frequency of agglutination of French articles to nouns (e.g. *lavi* [< *la vie*] 'life', rather than specifically 'the life').

Although probably genetically unrelated, Réu and IdF do share a small number of features not found in the New World French creoles, including the depalatalization of postalveolar fricatives, the use of *zot* (< *vous/les autres*) as both 2pl and 3pl, the pluralizer *bann* (< *bande*), and possibly also the predicate marker *i*.

References

Armand, Alain. 1987. Dictionnaire kréol rénioné-français. Saint-André: Éditions Océan.

Baggioni, Daniel. 1990. Dictionnaire créole réunionnais/français. Paris: l'Harmattan.

Baker, Philip. 1972. Kreol. A description of Mauritian Creole. London: C. Hurst.

Baker, Philip, and Chris Corne. 1982. Isle de France Creoles: affinities and origins. Ann Arbor: Karoma.

Baker, Philip, and Vinesh Hookoomsing. 1987. Diksyoner kreol morisyen. Paris: l'Harmattan.

Bollée, Annegret. 1977. Le créole français des Seychelles: Esquisse d'une grammaire, textes, vocabulaire. Tübingen: Max Niemeyer.

Carayol, Michel, and Robert Chaudenson. 1989. Atlas linguistique et ethnographique de la Réunion, 2 vols. Paris: Éditions du CNRS, 1984–1989.

Cellier, Pierre. 1985. Comparaison syntaxique du créole réunionnais et du français. Saint-Denis: Université de la Réunion.

Chaudenson, Robert. 1974. Le lexique du parler créole de la Réunion. Paris: Éditions Honoré Champion.

Corne, Chris. 1977. Seychelles Creole grammar. Tübingen: Günther Narr.

Offay, Danielle d', and Guy Lionnet. 1982. Dictionnaire créole seychellois-français. Hamburg: Helmut Buske.

Stein, Peter. 1982. Connaissance et emploi des langues à l'Île Maurice. Hamburg: Helmut Buske.

MIKAEL PARKVALL

Indian Traditional Grammar

It is to the Indian past that linguists of all countries and all theoretical persuasions still look for one of the crowning glories of intellect in linguistic science—to Panini's grammar of the Sanskrit language, written,

we think, some two and half millennia ago. The great grammarians of Sanskrit—Panini (phonetically, Pānini), and his followers Katyayana and Patañjali—created in their grammars of the Sanskrit language

linguistic masterpieces, as awe-inspiring in the realm of language as, in other esthetic spheres, the Mona Lisa, the works of Euclid, the Taj Mahal, or the works of Shakespeare.

Sanskrit was the language of ancient northern India and of the sacred writings of Hinduism. We believe that Indo-European (also called Indo-Aryan) invaders came into India from the northwest around 2500 BCE. The Dravidian family of languages was indigenous to India, and Sanskrit developed out of the interaction between the Dravidian substratum and the Indo-European adstratum. By about 350 BCE, Sanskrit had begun to diverge regionally, as all languages spoken over larger geographical areas do, given enough time. The Hindu grammarians broadened their interest from the Scriptures to the notion of 'correct' Sanskrit, making rules and lists of forms descriptive of the correct type of speech in every sphere: phonology, morphology, and syntax. Sanskrit was a part of every Hindu ritual, a sacred language, and it was essential that it be pronounced and inflected correctly to the last detail.

The culmination of this intense activity was the oldest treatise that has come down to us, the grammar of Panini. This grammar (the *Astadhyayi*), which dates from between 350 and 250 BCE, is one of the great monuments of human intelligence. It describes in detail every usage of its author's speech. Virtually no other language, even today, has been so completely described. Sanskrit became the official and literary language of all of Hindu India, even to the farthest reaches of the south. It ceased to be spoken as anyone's native language probably around the beginning of the Common Era, but it remained (as classical Latin remained in Europe and Hebrew in Judaism) the artificial medium for all writing on learned or religious topics, the language of the pundits.

Upper-class education in ancient India began with the acquisition, as a child, of the Sanskrit language. This was done by discussion and above all by memorizing Panini's grammar or one of the grammars based on it. Well after Sanskrit had ceased to be a widespread spoken language in India, classical Sanskrit literature was being written by men who had learned their Sanskrit from grammars and dictionaries, and teach-

ers, as we might learn Latin or Greek today. Panini was taken to be infallible: any deviations in Sanskrit compositions from the forms prescribed by Panini were errors.

One of the crowning achievements of Indian traditional grammar was phonetics. The Indian grammarians possessed a control of the technical details of phonetics that remained unmatched until well into the nineteenth century. Unlike most contemporary classifications of sounds, which use different features for vowels (high, low, back, rounded, unrounded, etc.) and consonants (voiced, voiceless, continuant, etc.), Panini's phonetic classification of the sounds of Sanskrit used a single grid for both consonants and vowels, as well as semivowels. Revised slightly and using modern phonetic terms, the phonemes of Sanskrit are as shown below:

Morphophonemics and morphology are described by means of rules, the exact nature and the interaction of which has been the subject of much debate. Typical—typical both of type and their intrinsic complexity—rules are:

a. *h* —> *ḍh,* if *h* is *saṃyogādi*, that is, if *h* is the first component of a consonant group when the suffix begins with a *jhal*-consonant (any consonant with the exception of a nasal or a semivowel). Example: *lih-dhi* —> *liḍh-dhi.*

b. A pair of rules (simultaneously ordered):
   ḍh-Deletion: ḍh —> zero / ___ ḍh
   Lengthening: i —> ī before a ḍh,
                      which is dropped

Certain aspects of Panini's treatment of morphology seem queer to us today; there is, for example, no specific recognition of the notion 'subject of a verb'. There is no case marking as such, even though Sanskrit has a richly developed case system. Perhaps the greatest accomplishment of the Indian grammarians in morphology is the notion that words are analyzable—can be broken down into roots and morphemes. Today, we think of the idea of 'root' as self-evident. It is not: Greek and Latin grammarians did not analyze words into roots and morphemes. Certain aspects of the Indian analysis of noun compounds are useful

Place of Articulation

Character of Sound	Glottal	Pharyngeal	Velar	Palatal	Labiovelar	Retroflex	Dental	Labial
Vowels		a ā		i í e ai	u ū o au	ṛ r̄	ı ī	
Stops			k kh g gh ŋ	č čhǰ ǰh ñ		ṭ ṭh d ḍh ṇ	t th d dh n	p ph b bh m
Semivowels				y		ṛ	ı	v
Continuants	h			ś				

508

today: *dvandva* = copulative (*bittersweet, flimflam, zigzag*), *bahuvrihi* ('much rice') = exocentric (*doorknob* is a kind of knob, *blackbird* is a kind of bird; these contrast with endocentric compounds such as *foul-up*, which is not a kind of *up*, or *turncoat*, which is not a kind of *coat*).

Panini's method has been summarized as follows. First, rules (*sutras*) are formalized. Second, metalinguistic elements (e.g. natural classes) are introduced for reasons of economy of description. Third, metarules (*paribhasas*), which explain how the rules work, are made explicit.

Linguistics in ancient India was the core of the intellectual and scientific tradition—it possessed an intellectual centrality and scholarly hegemony that beggar belief today. Linguistics was suffused with the light of sanctity, endowed with religious purpose. Panini and Patañjali and later Indian grammarians conceived their ministrations on behalf of the Sanskrit language as devotional. Their activity was more akin to a priesthood, a calling, than it was a profession or an academic discipline. Nevertheless, what the Indian grammarians produced over 2,000 years ago can hold its own with the best of linguistic theory today.

References

Allen, W. Sydney. 1953. Phonetics in ancient India. London: Oxford University Press.

Burrow, Thomas. 1965. The Sanskrit language. London: Faber and Faber.

Cardona, George. 1987. Sanskrit. The World's major languages, ed. by Bernard Comrie. New York: Oxford University Press.

Deshpande, M. 1993. Sanskrit and Prakrit, sociolinguistic issues. Delhi: Motilal Banarsidass.

Katre, S.M. 1954. Introduction to Indian textual criticism. Pune, Deccan College.

Rocher, Rosane. 1975. India. Current trends in linguistics, Vol. 13 (Historiography of linguistics), ed. by Thomas A. Sebeok. The Hague: Mouton.

Staal, J. F. 1972. A Reader on the Sanskrit grammarians. Cambridge, MA: MIT Press.

Whitney, William Dwight. 1889. Sanskrit grammar. Cambridge, MA: Harvard University Press. (often reprinted).

ROBERT D. KING

Indo-European 1: Overview

The Indo-European (IE) family is the most successful and widespread of the language families of the world. Seven out of the ten most widely spoken world languages are IE, and all of these seven belong either to its Romance, Slavic, Germanic, or Indo-Iranian branches. (These figures are taken from http://www.sil.org/ethnologue/top100.html (updated February 1999).) IE has spread over nearly all of Europe, plus large areas of Western Asia and the Indian subcontinent. Over the past few centuries, it has further spread to nearly the whole of the Americas, Australia, and New Zealand, plus most of Northern Asia, and small pockets in the Pacific and Southern Africa. This expansion has not been entirely one-sided: over the last 1,200 years, Hungarian and Turkic have gained at the expense of IE, specifically Slavic, Greek, Armenian, Iranian, and Tocharian.

IE is often considered from the point of view of its classical ancient languages (Latin, Greek, and Sanskrit), with supporting citations from Germanic and Slavic. These languages are very similar in overall structure, with nominal systems mostly distinguishing three numbers (singular, dual, plural) and three genders (masculine, feminine, and neuter), and case systems ranging from five cases (e.g. Greek, Germanic) to eight (Sanskrit), a fairly complex verbal system, characterized by distinctions of tense, voice, and mood, with some development of an aspectual system, synthesis rather than analysis, and a tendency to SOV (subject–object–verb) word order, etc. Most of the modern IE languages show greater or lesser degrees of reductions of these systems, with some languages showing remodelings. The most common reductions of these systems may be characterized by the following:

Grammatical Features

Gender: Reduction of original three-gender system to a two-gender system (mostly masculine/feminine, although sometimes common/neuter).

Number: Widespread loss of the dual.

Case system: General reduction of case, sometimes leading to the complete loss of the category, or increased syncretism.

Word order: In Europe, VSO and SVO orders have developed, with vestiges of SOV, while in Asian IE languages, SOV is still more widespread.

Most IE languages in Europe have also developed a system of definite and indefinite articles, with the definite article being more widespread than the indefinite one. They have also carried through the development of the verb 'have' more fully than the classical IE languages.

In broad outline, the following well-attested major branches are considered IE: Anatolian (the ancient languages of Asia Minor, e.g. Hittite, Luwian, Lydian, etc., attested between about 1700 BC and the fourth century BC), Indo-Iranian (e.g. Sanskrit (plus its descendants, e.g. Hindi, Bengali, Urdu, Marathi), attested from around 1000 BC; Avestan and Old Persian plus the modern Iranian languages, e.g. Persian, Kurdish, Pashto, also attested from around 1000 BC), Hellenic (Greek and its dialects, earliest attestations from 1400 BC), Italic (Latin (and its Romance descendants, e.g. French, Spanish, Italian), possibly plus other ancient languages of Italy, e.g. Oscan, Umbrian, attested from 500 BC), Celtic (e.g. extinct Continental Celtic languages, e.g. Gaulish, etc. attested in Roman times; modern Insular Celtic languages, e.g. Irish, Scottish Gaelic, Welsh, Breton attested from AD 300), Germanic (e.g. Gothic and other extinct East Germanic languages, English, German, Scandinavian languages attested from the second century AD), Baltic (e.g. Lithuanian, Latvian, attested from the sixteenth century), Slavic (e.g. Russian, Polish, Czech, Serbo-Croatian, Bulgarian, attested from the ninth century AD), Armenian (attested from the fifth century AD), Albanian (attested from the fourteenth century AD), and Tocharian (Central Asia, attested between the sixth and eighth centuries AD). Other extinct IE languages have been glimpsed in outline through early attestations, such as, e.g. Illyrian, Ligurian, Venetic, Pelasgian, Phrygian, and Thracian. In addition to the extinct IE languages noted above, the ranges of Hellenic, Baltic, Albanian, and Celtic have been severely restricted.

In this context, it should be noted that frequently family trees drawn for IE give distorted pictures, allotting nearly all the space to Germanic, Celtic, Italic, Indo-Iranian, and Balto-Slavic (nowadays sometimes also including Anatolian), while bunching Hellenic, Albanian, Armenian, and Tocharian together in a tiny slot. Despite their much smaller number of speakers, or lack thereof, it might be argued that together, the latter four groups encompass as much diversity as the former five.

The discovery of IE was a gradual process, with several stages and false trails. Thoughts by the Greeks and Romans on language were mostly confined to Greek and Latin, and later to the interrelationships between the two. For a long time, it was believed that Latin had descended from Greek, and this is reflected in the writing of authors such as Varro. This was later shown to be erroneous, although there is still a tradition of writing comparative grammars of Latin and Greek.

As the Middle Ages progressed, followed by the Renaissance, people became more aware of the linguistic interrelations within Europe, and the major families, and relationships between the coeval families, and the Classical languages. Medieval writers such as e.g. Giraldus Cambrensis and Otto of Bamberg noted several lexical items shared by Brythonic and Greek or Latin and Slavic, respectively. The general outlines of their linguistic interrelations slowly became clear during and immediately after the Renaissance. In 1599, Scaliger grouped the languages of Europe into four major and minor groupings. Scaliger's major groupings correspond to Romance, Greek, Germanic, and Slavic.

But the start of serious scholarship, and the realization that IE extended into Asia, is normally dated from Sir William Jones' famous address to the Royal Asiatic Society of Bengal in 1786, where he first conceptualized the idea of '*some common source*, which perhaps no longer exists' to explain similarities between Latin, Greek, and Sanskrit, the language of the Indian Brahmins, of which Europeans had been slowly becoming aware for about three centuries. Jones also suggested that 'both the *Gothick* (Germanic) and the *Celtick*' and the old '*Persian*' might also be derived from this 'common source'.

At first glance, Sanskrit appears even more archaic than Latin or Greek. For a while, it was even believed that Sanskrit was the ancestor of all IE languages. As the reconstruction of IE progressed, it was revealed that in some respects Sanskrit was less archaic than other IE languages (see below). Moreover, a later part of Jones' address deals with the modern languages of India, which he did not believe as descended from Sanskrit. In a lesser-known passage, he stated: 'and this analogy might induce us to believe, that the pure Hindi, whether of *Tartarian* or *Chaldean* origin, was primeval in Upper *India*, into which the Sanskrit was introduced by conquerors from other kingdoms in some very remote age...'; it was only later that it was discovered that Hindi stood in roughly the same relationship to Sanskrit as French to Latin.

Much research followed, and modifications to the theory came with time, but the basic edifice of IE remained. In the 1860s, Schleicher even felt confident enough to compose a fable in the reconstructed IE language. Schleicher's fable was later reworked by Hirt in

the 1920s, and by Lehmann in the late 1970s, but as the century progressed it became clear that IE was unamenable to this sort of analysis. The second half of the nineteenth century witnessed numerous apparent irregularities worked out. The formulations of the explanations for these irregularities were usually called 'laws', and this usage has stuck, although certain scholars were at pains to point out that they could not be equated to more universal laws, such as those of physics. These 'laws' were usually named after the scholars who formulated them.

Some of the major such 'laws', each representing a substantial step in the reconstruction of IE, and an advance in general linguistic theory, run as follows:

(1) Grimm's Law (1822) systematized the Germanic sound shift already noted by several scholars, where voiceless stops developed to voiceless fricatives, voiced stops to voiceless stops, and voiced aspirates stops to plain voiced stops (e.g. *t > th*, *tráyas/three*, *d > t*, e.g. *da a/ten*, *dh > d*, e.g. *vidhava/widow*; English forms contrasted with Sanskrit). Although many previous scholars had noted individual manifestations of this phenomenon, it is normally named Grimm's Law, as Jakob Grimm was the first to systematize it, grouping the voiceless stops (*tenues*), the voiced stops (*mediae*), and the voiced aspirates (*mediae aspiratae*) together in classes.

(2) Grassmann's Law (1863), which accounted for apparent irregularities in the distribution of voiced aspirates and voice stops, found in Greek and Sanskrit (sound changes could be spread over more than one syllable).

(3) Verner's Law (1876), which explained the role played by accent in apparent anomalies in the distribution of the reflexes of Grimm's Law (the role of accent in sound change).

(4) The Law of the Palatals (1870s), which is difficult to ascribe to any single scholar. It showed that the system of five short vowels attested in Latin and Greek (*a*, *e*, *o*, *i*, *u*) was actually more archaic than the system of three short vowels attested in Sanskrit (*a*, *i*, *u*), previously seen as the original system.

Shortly after all these laws had been formulated, the so-called Neogrammarian school emerged. They proposed that there were no exceptions to sound laws as a working hypothesis and compiled numerous comparative grammars, both of IE (Brugmann, Delbrück, Osthoff) and of individual IE languages (e.g. Leskien on Old Church Slavonic). Mainly centered in Germany, they were very influential on IE studies, and the vast compilations of their material are still consulted today, although much has been superseded. IE reconstructed at this stage in the development of the field is sometimes called 'Brugmannian IE', after one of the most prominent Neogrammarians.

Attempts were also made to delineate subgroupings within IE. The major one normally proposed, and elevated to considerable importance in the late nineteenth century, was the *centum/satem* split (from the words for 'hundred' in Latin and Avestan, respectively). In certain forms, the *centum* languages retained an original velar, while the *satem* languages developed the velars into a palatal or sibilant. Throughout most of this century, however, this postulated *centum/satem* distinction has steadily been losing importance in IE studies.

One ongoing area of IE studies is the attempt to demarcate early dialectal divisions within IE. The only universally accepted one, however, is Indo-Iranian. It was also once popularly believed that Italic and Celtic were closely related, based on a few common items of vocabulary and a few shared grammatical innovations and archaisms. Certain scholars have seen a European IE grouping of languages, mainly based on shared lexical features. There is an ongoing controversy about whether Baltic and Slavic are two separate branches of IE that have come to resemble each other through convergence, or whether an intermediate Balto-Slavic grouping may be reconstructed.

In the early years of this century, new discoveries and decipherments shed fresh light on IE, while at the same time making the picture more complicated. The first was the discovery in Western China in 1900 of manuscripts in a previously unknown IE language, Tocharian, split up into two heavily differentiated dialects, conventionally called Tocharian A and Tocharian B. The structure of Tocharian, while recognizably IE, differed in several important aspects from the Brugmannian system.

A few years later, a massive corpus of material written in cuneiform was unearthed in Anatolia. The decipherment and linguistic interpretation of the material took some time, but finally it was recognized that here was another new IE language, identified with the language of the Hittites of the Old Testament. While definitely IE, the structure of the newly interpreted language was quite unlike what had been expected: in many ways, it was closer to modern IE languages, including a fully developed verb 'have'. The verb morphology was far less complex than Greek or Sanskrit, and this caused some scholars to give serious reconsideration to the reconstruction of IE morphology. As research progressed, it was discovered that Hittite was only one member of a whole family of IE languages, normally referred to as Anatolian (see above). The most important

contribution from Anatolian, however, was its apparent confirmation of the 'laryngeal theory'. In the late 1870s, speculations about the IE vowel system had led de Saussure to postulate the existence of otherwise unattested segments, which were later reconstructed as laryngeals. These segments were actually attested in Hittite. Not all linguists accept the laryngeal theory, although the discovery and interpretation of the Hittite material have given it a tremendous accession of support.

The discovery and interpretation of Tocharian and Hittite material also caused the *centum/satem* theory some problems, as both of them showed *centum* reflexes, despite being located to the east. It has taken time for the results of research into Hittite and Tocharian to be properly integrated into mainstream IE studies, and perhaps, the process is not yet complete.

IE continues to be subject to modification and reinterpretation. It has even been quipped that 'no language has changed as much in the past fifty years as IE'. Areas where there has been much discussion recently include the glottalic theory (a reinterpretation of the nature of the IE stops), ergativity (subject–object relations), and the degree of morphological complexity of the protolanguage.

References

Beekes, Robert. 1995. Comparative Indo-European linguistics. Amsterdam: John Benjamins.

Birnbaum, Henrik, and Jaan Puhvel (eds.) 1966. Ancient Indo-European dialects. Berkeley: University of California Press.

Collinge, N.E. 1985. The laws of Indo-European. Amsterdam: John Benjamins.

Ivanov, V.V., and Gamkrelidze, T.V. 1995. Indo-European and the Indo-Europeans, translated by Johanna Nichols. New York: Mouton de Gruyter.

Lockwood, William.B. 1969. Indo-European philology. London: Hutchinson.

Ramat, Paolo, and Anna Giacalone Ramat (eds.) 1998. The Indo-European languages. London and New York: Routledge.

The website http://www.dla.utexas.edu/depts/lrc/iedocctr/ie-links.html at the IE Documentation Center at the University of Texas at Austin is also an excellent resource.

ROBERT ORR

See also **Indo-European 2: Germanic Languages; Indo-European 3: Indo-Iranian Languages; Indo-European 4: Romance; Indo-European 5: Slavic; Jones, Sir William**

Indo-European 2: Germanic Languages

The Germanic languages are spoken by more than 450 million native speakers and display a very wide geographical distribution, partly due to the current status of English as the world's most important international language.

It is generally assumed that by the first century BCE, Germanic people speaking a fairly uniform language were living somewhere in southern Scandinavia, the Danish Isles, and northern Germany, on both sides of the North and Baltic Seas, and they started to expand rapidly during the Roman period, primarily to the detriment of Celtic languages. In time, there developed the so-called West, East, and North Germanic dialects. The West Germanic tribes settled in the lands between the Elbe and Oder rivers and to the immediate south and west of that area (about the area of modern Germany and the Benelux states), and it is here that the German language gradually evolved. The East Germanic tribes settled in the Baltic area east of the Oder River (the area of modern Poland), but their languages have long since become extinct. In Scandinavia, the North Germanic tribes spoke a language we now call Old Norse, the ancestor of the modern Scandinavian languages. In the fifth century CE, three West Germanic tribes, the Angles, Saxons, and Jutes, crossed the North Sea into Britain, bringing with them a language that would later be known as English. And in the ninth century, Old Norse was carried far westward to Iceland. These wanderings left Europe dotted with Germanic dialects.

The Germanic languages offer extensive and homogeneous documentation from a relatively early period. The oldest monuments of the Germanic languages are more than a hundred Scandinavian inscriptions in a special alphabet called Runic. They date from the third or fourth century CE and are representative of the North, or Scandinavian, group of Germanic languages.

The Germanic languages have very clear-cut and readily identifiable features distinguishing them from the other Indo-European stock. The best known is described by the so-called Grimm's Law (or the First Sound Shift), referring to a systematic shift in the Indo-European consonant system during the development of the Germanic ancestor language. Besides, the

Indo-European verbal system was simplified. The Germanic languages developed a preterite tense (called weak or regular) with a dental suffix, -d or -t (e.g. *fish, fished,* etc.). Germanic languages thus have two types of verbs: weak (regular) and strong (irregular). Strong verbs indicate tense by an internal vowel change (e.g. *ring, rang, rung*).

In terms of vocabulary, Germanic has a number of unique words that are not found in other Indo-European languages. These words may have been lost in the other Indo-European languages, borrowed from non-Indo-European languages, or perhaps coined in Germanic.

Numerous groupings of the Germanic languages have been proposed during the last centuries, beginning effectively with the grammar of Jacob Grimm in 1819. Traditionally, the Germanic languages are divided into East Germanic (with Gothic as its most important member), North, and West (sometimes 'South Germanic') subgroups.

Besides the living Germanic languages described below, there are several extinct languages such as Bastarnae, Burgundian, Frankish, Gothic, Herulian, Lombardic, Norn, Rugian, Scirian, and Vandalic.

East Germanic

The Goths were the first Germanic tribe to leave their Germanic homeland (presumably around 100 BCE) and due to this early migration their language developed differently from those of the other Germanic peoples.

Gothic offers the oldest literary record of Germanic languages: a translation of a Greek bible into Gothic by Bishop Wulfila in the fourth century CE. It is an extremely regular and consistent translation and includes a large part of the New Testament and some parts from The Old Testament.

Gothic has played an important part in the reconstruction of the Indo-European and Germanic ancestor language. Its structure is reminiscent of the other old Indo-European languages, with complex inflections of noun, verb, and adjective. It was heavily influenced by Greek in Wulfila's translation of the Bible and this influence is evident with respect to the lexicon, word order and sentence structure.

North Germanic

North Germanic is commonly divided into two subgroups gradually splitting up from the Common Scandinavian toward the end of the Viking Age (800–1050): the eastern one, including Swedish, Danish, and Gutnish (on the Baltic island of Gotland), and the western one, including Norwegian, Faroese,

and Icelandic. The oldest North Germanic language is referred to as Old Icelandic, or Old Norse, which suffered a rapid change after 700, due to the restless movements of the Viking Age. However, Old Norse is the language of an extraordinarily rich and varied literature inspired by the traditions and manners of the incomparable Viking Age.

Danish

Danish, the mother tongue of over 5 million people, is the official language of the kingdom of Denmark (including Denmark, Greenland, and the Faroe Islands). The modern language developed on the basis of the written language of Reformation influenced by the seventeenth to eighteenth centuries language of Copenhagen, the economic and cultural center of the emerging nation state. Of all Scandinavian languages, Danish has differentiated the furthest from the Common Scandinavian ancestor.

Swedish

Swedish is the official language of Sweden and it is spoken by 8.5 million Swedish and 300,000 Finlanders. It was first recorded in about 2,000 runic inscriptions from the eleventh and twelfth centuries. The modern language developed in the Mälar-Uppland region, where Stockholm and Uppsala have been raised as the main centers of government and learning since the Middle Ages.

Norwegian

Norwegian is the official language of Norway and the mother tongue of over 4 million people. It has two varieties, Neo-Norwegian (*nynorsk*) and Dano-Norwegian (*bokmål*), both official languages used by national and local officials. Nynorsk is more prevalent in rural areas, Bokmål in the cities. Bokmål is spoken by approximately 90% of the population. These two modern standards have a historical cause: during 1380–1814, Danish was the written language of Norway, while most Norwegians used their local dialects and pronounced Danish with their own Norwegian sounds. Neo-Norwegian is based on local dialects and was officially recognized in 1885. Dano-Norwegian, or 'book language', is the first language of the majority of the population.

Faroese

Faroese is spoken by 47,000 inhabitants in the Faroe Islands, where it is also the official language (together with Danish). It is a very archaic variety of North

Germanic isolated during the Viking expansion and it developed from the language of the Norwegians who colonized the islands in the early 800s. Faroese, a descendant of Old Norse, has few speakers, but a lively national literature for nearly 200 years. It has only come into relative prominence in modern times. Structurally, Faroese corresponds more to Middle Norwegian than to Icelandic.

Icelandic

Icelandic is the official language of Iceland where it has been spoken ever since the country was settled over a thousand years ago. It is the mother tongue of 250,000 people. Icelandic is the most conservative North Germanic language, due to its geographical isolation and the fierce purism of its speakers, i.e. no other Germanic language has remained so close to the ancestral form. It has never shown any real tendency to split into dialects. It has a rich literature dating from the thirteenth century.

West Germanic

West Germanic is considered to be the least unified of the three major groups. The records are relatively late and very uneven in chronology and content. Each of the individual languages is highly inconsistent and dispersed into various dialects. However, two main branches can be identified: the English group, which is usually considered with Frisian as Anglo-Frisian, and the German group, including High and Low German.

English

English is an official language of Britain, Canada, the United States, Australia, New Zealand, Ireland, the Bahamas, Jamaica, Barbados, Grenada, Trinidad and Tobago, St. Lucia, Belize, Guyana, Botswana, Cameroon, Dominica, Gambia, Ghana, Gibraltar, Lesotho, Liberia, Malawi, Nigeria, the Seychelles, Sierra Leone, Singapore, South Africa, Swaziland, Uganda, Vanuatu, Western Samoa, Zambia, and Zimbabwe; it is the associate official language of India. It is the native language of more than 300 million people and the second language of another 300 million. English is also the universal language of trade, communications, science, and transportation and the most frequently acquired second language in the world.

English is characterized by a very large vocabulary, nonphonetic spelling, an almost total lack of inflection (most plurals of nouns are indicated), a syntax almost totally dependent on word order, and a very complicated periphrastic verb system.

Frisian

Frisian is spoken in northern Holland (the province of Friesland and some adjacent islands), and also in Germany (the low coastal areas). It is one of the two official languages of the Netherlands. Of all Germanic languages, Frisian is most closely related to English. The oldest records are a series of legal documents dating from the tenth to the sixteenth centuries. Structurally, the language is very similar to Dutch.

Dutch

Dutch is the official language of the Netherlands, Suriname, and the Netherlands Antilles, and Flemish is the official language of Belgium. The language is also spoken in small parts of Germany and France, Indonesia and elsewhere in Southeast Asia. It is native to some 14 million people. The Dutch spoken in Belgium is called Flemish, although there is practically no difference from Dutch in the Netherlands. Standard Dutch is based on the dialect of the Amsterdam region after it became the capital of an independent nation.

Afrikaans

Afrikaans, sometimes considered a creole of Dutch, is an official language of South Africa and Namibia. It is spoken by 10 million people. Afrikaans developed in the seventeenth century from the Dutch brought to South Africa by the first settlers from Holland. Originally, it was a popular dialect composed of Dutch with considerable borrowings from aboriginal languages of Africa, especially Malay (spoken by the slaves in the seventeenth century).

German

German is an official language in Germany, Austria, and Switzerland; it is also spoken in eastern France (Alsace and Lorraine), Belgium, Luxembourg, Liechtenstein, pockets of the United States, parts of Eastern Europe, and in former German colonies (Namibia, Congo, Cameroon). German is also used as a second business language for Continental Europe and Scandinavia.

Following the so-called 'High German Consonant Shift', the dialects of German are divided into Low German (*Niederdeutsch, Plattdeutsch*) and High German (*Hochdeutsch*). Low German is spoken in the lowlands of northern Germany. The term 'Low German' is essentially a geographic term referring to the coastal, or lowland, area of the German region, as opposed to the High German area. Modern Standard

German developed mainly from the language of the late medieval chancery language of the court and Saxony and the East Central dialect area of Dresden.

Yiddish

Yiddish (also called Judeo-German) is a High German language spoken throughout northern continental Europe (especially Germany, Poland, Lithuania, and Russia) and North America by the Jewry. Originally, it was the language spoken by East European Jews, who migrated from Germany in the fourteenth to sixteenth centuries. There are two varieties of Yiddish: an eastern one developed under Slavic influence and developing a rich literature in a highly standardized language, and a western one developed under German influence. Yiddish literary records go back to about the thirteenth century. Like all Jewish languages, it is written in Hebrew characters read from right to left.

Pennsylvania German

Pennsylvania German, also known as 'Pennsylvania Dutch' or simply 'Dutch', is spoken by 300,000 native speakers, descendants of German colonists, mainly in the United States. It is the vehicle of a typical folk literature and its speakers are all bilingual, as the official language is English.

References

Abraham, Werner, Wim Kosmeijer, and Eric Reuland (eds.) 1991. Issues in Germanic syntax. Berlin, New York: Mouton de Gruyter.

Braune, W., and E. Ebbinghaus. 1981. Gotische grammatik (Gothic grammar), 19th edition. Halle: Niemeyer.

Comrie, B (ed.) 1987. The World's major languages. London: Routledge.

Haugen, E. 1982. The Scandinavian languages. A comparative historical survey. Tübingen: Niemeyer.

Helbig, G., and J. Buscha. 1991. Deutsche Grammatik (German grammar), 14th edition. Berlin: Langenscheidt.

Lippi-Green, Rosina L (ed.) 1992. Recent developments in Germanic linguistics. Amsterdam, Philadelphia: J. Benjamins.

Lippi-Green, Rosina L., and Joseph C. Salmons (eds.) 1996. Germanic linguistics. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Rauch, Irmengard, Gerald F. Carr, and Robert L. Kyes (eds.) 1992. On Germanic linguistics: issues and methods. Berlin, New York: Mouton de Gruyter.

Russ, C.V.J. (ed.) 1990. The dialects of Modern German. A linguistic survey. London: Routledge.

Stephens, M. 1976. Linguistic minorities in Western Europe. Llandysul: Gomer Press.

Vikør, L. 1993. The Nordic languages: their status and interrelations. Oslo: Novus.

LAURA DANILIUC

Indo-European 3: Indo-Iranian Languages

Indo-Iranian languages are the easternmost subfamily of the I[ndo]-E[uropean]family, spoken by almost a billion people, chiefly in Iran, Afghanistan, Tajikistan, Bangladesh, India, Nepal, Pakistan, and Sri Lanka. The languages of this subfamily are among the oldest of the Indo-European group, and are well represented among the oldest records of Indo-European languages. However, there is debate about the relation of these languages to the Hittite languages of Anatolia. Some references believe them to have originated around modern Afghanistan, and some others name Turkistan. This branch consists of three groups of languages, the *Iranian*, the *Indo-Aryan* (or *Indic*), and the newly added *Dardic* (also known as *Kafir* or *Nuristani* languages), preserved in the Himalayan Mountains.

The Indo-Iranian group is remarkable, both in terms of its history and in terms of the number of its speakers. Two of the ten most widely spoken languages on earth today, Hindi and Bengali, belong to the Indic branch of Indo-Iranian. Indo-Iranian languages also have a very old history. The Indic and the Iranian languages have passed through three historical periods, namely old, middle, and new.

Iranian

The Iranian languages are mainly spoken in today's Iran, Afghanistan, Tajikistan, and some parts of Pakistan, Turkey, Iraq, and former Soviet countries. It is estimated that about 95 million people speak in these languages.

Old Iranian

Historically, the oldest Iranian languages of which there are records available are Old Persian and Avestan, both highly inflected languages. Old Persian has survived in cuneiform inscriptions from the time of the Achaemenid kings, who ruled ancient Persia during

the sixth to fourth centuries BC. Avestan is the language in which *Avesta*, or the sacred text of the Zoroastrian religion was composed. The *Avesta* probably dates from about the seventh to the fifth centuries BC, but apparently was handed down orally and was not recorded in writing until much later. Both of these languages are extinct, although Avestan is still formally in use as the religious language of the Zoroastrians.

Middle Iranian

The Middle Iranian period, dating from the third century BC to the ninth or tenth century AD, is characterized by considerable grammatical simplification, as in the reduced inflection of the noun and verb. Among the languages surviving in written records that fall within this period are Parthian, Middle Persian, Khwarazmian, Sogdian, and Saka.

New Iranian

The new Iranian languages are those languages used in the region after the emergence of Islam in the region (ninth or tenth century), and represent more phonetic and grammatical simplifications. For example, many case endings have dropped, and have been substituted by adpositions. The new Iranian languages are considerably more in number in comparison with the middle Iranian languages. What is interesting about these languages is that most of them do not have a direct and specified ancestor, and only Persian and Yaghnobi can be regarded as those having direct ancestors, Middle Persian and Sogdian, respectively.

The new Iranian languages are now divided into 'eastern' and 'western' subgroups. New west Iranian languages are basically spoken in today's Iran, and include Persian (the widest used Iranian language, with Iranian, Afghan, and Tajik variations), Kurdish, Baluchi, Mazandarani, Lori, Gilaki, Tāti, Tāleši, Lārestāni, Bashkardi, and some others. There are languages, of course, spoken beyond this border, and yet considered to be west Iranian, like Baluchi (which is also spoken in Pakistan), Zazaki (spoken in Turkey), and Kumzari (spoken in Musandam peninsula in Oman). Most of these languages do not have a writing system, and make use of Persian alphabet. Iranian Persian is also written this way, but the Tajiks make use of Cyrillic alphabet.

Middle Iranian languages have been ergative languages, and this feature has still been retained in many modern Iranian tongues. Persian, Lori, Mazandarani, and Gilaki are, of course, among the nominative-accusative Iranian languages.

The east Iranian languages are spoken in Pamir region; Afghanistan, Tajikistan, and to a much lesser extent, central Asia. They are numerous and scattered,

and some of them are greatly in danger of extinction. Some of the best-known languages of this group are Pashto, Wakhi, Ishkashmi, Yoghnobi, Yazgulami, Shughni, Parachi, Bartangi, Rushani, and Sarikoli (the easternmost Iranian language).

Indic

The Indic languages form the largest group of the Indo-Iranian subfamily. This group contains about thirty languages, some of which with numerous dialects. About 700 million people speak in these languages, of which the most widely used languages are Hindi-Urdu, Bengali, Punjabi, and Marathi (to these languages, we should add the Gypsy or Romano languages, which are scattered in various parts of Europe, as well as in the Indian subcontinent). Having voiced aspirated plosives like [bh], as well as retroflex sounds in the phonological system, are regarded as the two main characteristics of these languages. Like the Iranian languages, Indic languages have a long history, and may be studied within Old, Middle, and Modern Periods.

Old and Middle Indic

The oldest written form of the Indic group is Vedic, which is the language of the religious hymns of Veda. The Veda represents an oral tradition preserved by the ancestors of Hindu priests. Vedic has been described as the parent language of Sanskrit which by the fourth century BC had become the sacred and literary language of the Hindus of India, and its classical form was in use (at least for literature) until c. AD 1100. Sanskrit, as the most important source for all modern Indic languages, has survived to this day as a liturgical language in India. It is written in Devanagari, a development of the Brahmi script. Most Indic languages are written in some modified form of the Devanagari alphabet.

Prakrit (or 'natural' as opposed to Sanskrit 'polished'), which represents the Middle Indo-Iranian, refers to the state of the language between the third century BC and the fourth century AD. The best-known language of this era is Pali, in which the Buddhist Ašoka scriptures were originally written.

Modern Indic

Modern Indic languages have undergone extensive changes in the course of their evolution. Although the vocabulary of many Indic languages derives primarily from Sanskrit, Muslim influence over the centuries has added loan words from Arabic and Persian to some Indic languages such as Urdu and Sindhi. Besides phonemic and vocabulary changes, the simplification of the inflection of nouns and verbs is quite considerable. Cases of the nouns have been reduced from eight

to two, and are mainly represented by prepositions, rather than the earlier case markers.

Modern Indic languages are so many; thus their classification is also diverse. They can be classified as 'Northwest Indic languages' (like Punjabi and Sindhi); 'Central Indic languages' (like Hindi and Urdu); East Indic languages (like Assamese, Bengali, and Oriya); the West Indic language (Gujarati); South Indic languages (like Marathi and Singhalese); and the northern or Pahari languages.

As to the population of the speakers, Hindi, now the national language of India, stands out, and is understood by about half a billion people, mostly living in central India. Urdu is the other language, so similar to Hindi (but recorded by Persian alphabet), spoken mainly in Pakistan and India. Punjabi, spoken in Northwest India and Pakistan, is close to the Western Hindi dialect and is written in an alphabet based on the Devanagari script. Sindhi, the native language of people in Southeast Pakistan and West India, is the other language of this group, recorded in a modification of a Persian script by Muslims, although Hindus use a variety of the Devanagari alphabet as well.

Of the East Indic languages, Assamese is noteworthy. Although reminiscent of Sanskrit in vocabulary, it is allied grammatically to Bengali, the national language of Bangladesh. Oriya, the language of about 32 million persons, chiefly in the Indian state of Orissa, is closer to Sanskrit phonetically and lexically than any other modern Indic tongue. The leading West Indic language is Gujarati, spoken chiefly in the states of Gujarat and Maharashtra in India.

Of the South Indic languages, Marathi is spoken in the Indian state of Maharashtra. The other important South Indic language is Singhalese, the language spoken by people of Sri Lanka. This language is geographically separated from the other Indic languages of North and Central India by an intervening region in South India, in which the Dravidian languages are spoken. Thus, it is greatly influenced by the Dravidian languages nearby. The Pahari languages (or dialects, which are so many in number) are spoken by the people in Nepal and North India. Pahari has two main dialects, Garhwali and Kumaoni, and Western Pahari.

Dardic

There is no general agreement among the linguists on the status of Dardic group. Many references believe them to be a subgroup of the Indic, but there are firm evidences that classify them as a separate category. Dardic, or Pisacha, languages are spoken in Afghanistan. These languages share certain distinctive phonetic characteristics, feature the use of pronominal suffixes with various verb forms, and include in their vocabularies a number of words that among the languages of India are usually encountered only in Vedic Sanskrit. Kashmiri is the sole Dardic language that both has a literature and is recognized in the Indian constitution of 1950. It is written in Persian letters by Muslims, whereas Hindus use a script similar to the Devanagari alphabet.

References

Boyle, Grammar of modern Persian (1966); C.P. Masica, The Indo-Aryan languages (1989).

Campbell, George L. 1991. Compendium of the World's languages. London: Routledge.

Chatterji, S.K., Indo-Aryan and Hindi (2nd edition 1960); A.M. Ghatage, Historical Linguistics and Indo-Aryan Languages (1962); J. Bloch, Indo-Aryan, from the Vedas to modern times (revised edition, tr. 1965); T. Burrow, The Sanskrit language (2nd edition 1965).

Crystal, J.A., David. 1987. The Cambridge encyclopedia of language. 2nd edition. New York: Cambridge University Press.

Grimes, Barbara F. (ed.) Ethnologue. Summer Institute of Linguistics, ongoing.

Lyovin, Anatole V. 1997. An introduction to the languages of the World. Oxford: Oxford University Press.

Ruhlen, Merritt. 1987. A guide to the World's languages: classification, Vol. 1. Stanford, CA: Stanford University Press.

Schmitt, R. 1992. Compendium linguarum Iranicarum. Wiesbaden: Dr. Ludwig Reichert.

BEHROOZ MAHMOODI-BAKHTIARI

Indo-European 4: Romance

Spoken throughout the Roman Empire, around the Mediterranean Sea, and even beyond, Latin, an Indo-European language, was the starting point of a dialectal differentiation that gave rise to Romance languages. More precisely, it was its late form by the end of the Empire (fourth century CE), which was also called 'Vulgar Latin'. It is difficult to pinpoint the moment when Romance languages first appeared, however,

since these idioms are extensions of the Latin language. Today, Romance languages in Europe extend from the Atlantic to the Black Sea. From west to east, the following national languages can be found: Portuguese, Galician, Spanish, Catalan, Occitan, French, Rhaeto-Roman, Italian, Sardinian, and Romanian, making up what is known as *Romania*. The international expansion of these languages must also be taken into account. Linguistically, they are divided into two entities: western Romania (Italo-Roman, Ibero-Roman, Gallo-Roman, Rhaeto-Roman) and eastern Romania (Daco-Roman).

Italian

Italy is the epicenter of the Italian language: 95% of its population speaks it. Dialectical segmentation, however, is very important there. Three key subgroups can be identified: in the north, Septentrional Italian with Gallo-Italic (Piemontic, Lombardic, Genoese) and Venetian, Tuscan in the center (but also in Sardinia and in Corsica), and the Central-Meridional (ranging from Roman [Rome] to Calabrian and Sicilian [Calabria, Sicily]). This fragmentation slowed the standardization of the Italian language: Venetian first played this part, but it was Tuscan (Florence) that triumphed from the nineteenth century onward, even though all the regional dialects still remain quite lively today. Italian (Venetian) also remained a vehicular language up until the end of the nineteenth century. Although it still retains traces of its past expansion in Africa (Somalia, Ethiopia, Erythrea), Italian now radiates mostly through the emigration of its speakers. There are thus large communities of Italians in America (6 million), Australia (1 million), or Europe (2 million, mainly in Belgium). Overall, this Romance language is spoken by about 66 million people, 57 million of whom are in Italy.

Sardinian

Still being used in Sardinia by 1.2 million people, Sardinian language has a conservative character compared to other Romance languages. Its lexicon is archaic, except for late borrowings from Catalan, Castilian, and Tuscan languages. There are two subgroups: Logudorian (in the north) and Campidanian (in the south). The vast majority of Sardinian speakers are bilingual Sardinian–Italian, although its recent status as an autonomous region is giving it new strength.

Spanish

The Spanish Language has three main dialects: Astur-Leonese from Asturias to Salamanca, Aragonese in the southern Pyrenees, and Castilian (ancient Castile), which, starting with Alfonso X El Sabio (thirteenth century), became dominant and the basis for modern Spanish language, even though today there are attempts to rehabilitate Aragonese. Spanish is the Romance language that has grown the most: it can be heard in Israel where there is a large Judeo-Spanish colony (about 350,000 speakers) and also almost all over Latin America since the conquest of the Aztec and Inca Empires in the sixteenth century, when it became the official language of all these countries. Its expansion continues in US States contiguous to Mexico (Arizona, New Mexico, California, Texas, etc., with at least 20 million speakers). The only noteworthy Creole that developed from contact with Spanish is that of the Dutch Antilles, *Papiamento*, which is spoken by 200,000 people and is an official language. Spanish is not widespread in Africa and only 2 million Filipinos use it in Asia. Overall, 300 million people are Hispanophones.

Portuguese

The formation of modern Portuguese comes mostly out of Gallego-Portuguese, which used to link Portuguese and Galician dialects. Political independence allowed Portuguese to become a true Romance language. It is the result of the harmonious fusion of the different dialects from the provinces that make up the country: dialects from Coimbra, Lisbon, Beira, and Estramadure. The existence of Azorean and Madeiran, however, can be noted as their differences remain somewhat stronger due to their insular character and late colonization (fifteenth century). As with Spanish, but to a lesser degree, Portuguese lexicon was influenced by Arabic. On the other hand, it has imported more words from its (former) African colonies and its (ancient) Asian trading posts. Furthermore, it can be said that the notable differences found in the Portuguese spoken in Brazil are contributing to the creation of a new Romance language: Brazilian. The differences are most obvious in the lexicon, with borrowings from indigenous languages (Tupi) and from the languages of the former African slaves (Bantu), and in its syntax, which is at times archaizing. Brazilians now make up the largest Lusophone group in the world, with 150 million speakers. Further, Portugal's former African colonies have all retained Portuguese as the official language, but only 5% of their population (i.e. 2.5 million people) use it. Native people in these countries have often modified it, giving way to Creoles, such as Capeverdean Creole (Cape Verde), which is spoken by 230,000 people. Lastly, the Portuguese Diaspora can be found in France as well as in the United States of America (Massachusetts). All in all, it can be estimated that Portuguese is now

spoken by at least 170 million people all over the world, with the great majority of speakers being from outside its country of origin.

Galician

Geographically, Galician is located within a Spanish Province. Yet, it is a dialect from the Portuguese family and can be found in northern Portugal also. Its location made it lose its contact with Portuguese very early on. In fact, it mixes both Spanish and Portuguese characteristics. It began to reappear at the beginning of the twentieth century and is today spoken by about 2.5 million people (or 90% of Galicians), but its use is strongly declining in cities. Galician has, within its Spanish Province, the status of an official language.

Catalan

Its resemblance with Occitan leads one to consider Catalan as part of the Gallo-Roman group rather than Ibero-Roman. Its mixed lexicon, however, shows that this language is more like a 'bridge' between the two groups. Its strength is that it is an international idiom, since this language can be found in Spain (Catalonia, Baleares), as well as in France (Roussillon), Andorra, or Italy (Sardinia). It does not show any strong dialectical segmentation. In Spain, it counts six million speakers and served as the official language of Catalonia between 1931 and 1939. It has regained this status since 1979. Although it has the status of official language in Andorra, it does not have any real status in France, nor in Sardinia, where less than 20,000 people use it.

French

French is the one Romance language that has evolved the most compared to the original Latin. French language is located in the northern half of today's France (except for Brittany and Alsace), as well as in southern Belgium (Wallonia). It is part of the Langue d'Oil dialects, which were thus named in opposition to the Langue d'Oc languages (Occitan). These dialects are many. Some remain in use, such as Norman (Normandy) and Gallo (west Brittany). Walloon has even been recognized as a regional language in Belgium since 1990. Others have fully disappeared, such as Champenois or Berrichon, or are hardly surviving, such as Franco-Provençal (100,000 speakers in France and Italy). Each has its own set of particularities that would be difficult to detail herewith.

The expansion of French is multiple. It first grew within the country, with the progressive eviction of all other dialects (be they Gallo-Roman or not) as early as the eleventh century; it then spread across its borders (conquest of England in particular), and beyond (Lebanon). In the eighteenth century, at a time when France excelled, Francization reached Flanders and Russia. Later, the colonization of America and Africa (Maghreb, Black Africa) gave rise to Creole languages in America (Louisiana), Caribbean (Haiti, Antilles, Guyana), and some parts of the Indian Ocean (Madagascar, Reunion, Seychelles) as well as to the Québecker dialect (the official language of Québec) and Acadian (New-Brunswick). There are 70 million francophone speakers in Europe, 72 million in African countries (Senegal-Djibouti axle), and 45 million in the Maghreb. To this, the Caribbean (750,000), the Indian Ocean (600,000), and Lebanon must be added. There are about seven million Francophones in Québec and 280,000 more across the other Canadian Provinces. The Acadians in Louisiana number about 400,000 people to which 50,000 more speakers living in the eastern part of Louisiana and in Texas must be added. Thus, French is actually spoken in at least 38 states and the total number of French speakers is estimated to be a little under 200 million, with only 30% residing in Europe.

Occitan

Occitan dialects can be found in the south along a line that splits France from Bordeaux to Briançon. There are three subgroups: Septentrional Occitan (Limousin, Auvergnat, etc.), Southern Occitan (Languedocian, west of the Rhone, and Provencal in the east), and Gascon (in Aquitaine). Occitan is characterized by different phonetic and morpho-syntatic traits. Gascon, however, through its relation to Ibero-Roman, shows originalities to the point that it could be considered a Romance language in itself. Literary and judicial language until the Albigensian Crusade (1209–1229) and the annexation of its lands by the Crown, the Occitan then experienced a decline. It would not reemerge until the nineteenth century with the Felibrige Movement and F. Mistral. Since the second half of the twentieth century, autonomist movements have claimed it as a regional language. The number of speakers who master Occitan nowadays can be reasonably estimated at about 10 million.

Rhaeto-Roman

There is no real dialectical unity in this subgroup. Its specificity is mostly historical. Linguistically, it is hinging on both Romania, based on its lexicon in particular, but globally, it is close to Gallo-Roman. It is made up of three dialects, geographically fragmented: Romansch, spoken in Switzerland (Canton of Grison),

Ladin in Italy (Tyrolean Alps, north Trentino), and the Friulan also in Italy (Province of Udine). Wedged between Bavarian and Alemannic in the north (Germanic languages), and Lombard in the south (Italian subgroup), this Romance language has frittered away as the centuries went by. Today, there are only 36,000 Romansch speakers, 30,000 Ladinophones, and about 600,000 Friulan speakers. Depending on their geographical location, most are bilingual (German in the north, Italian in the south). And even though Romansch has been one of the national languages of Switzerland since 1938, assimilation seems irreversible.

Romanian

One characteristic of the Romanian language is that it is very conservative on some points while at the same-time being very innovative on others. The influence of Slavic and Byzantine cultures explains its originalities. The Romanian lexicon is under Slavic influence, as well as under Greek, Hungarian, and even modern French influence. In the eighteenth century, Slavonic stopped being the official language. Influenced by the ideas of the French Revolution and by an awareness of the Latin origins of the language, modern Romanian appeared at the beginning of the nineteenth century. The Latin alphabet became prominent in Romania beginning in 1860, but had to be adapted. The Cyrillic alphabet, however, was still used in Moldova up to the dissolution of the USSR. Romanian language is subdivided into four groups: Daco-Romanian (found in today's Romania and official language and in Moldova), Istro-Romanian (north of Thessalia), Macedo-Romanian (or Vlach, dispersed in Greece and Macedonia), and Megleno-Romanian (south of Bulgaria and north of Thessalonica). Istro-Romanian is spoken by only about a 1,000 people and is an endangered language. Macedo-Romanian is spoken by about 100,000 people and Megleno-Romanian, the dialect of a Muslim community, by 15,000 people. The unified Romanian language (Daco-Romanian) counts overall 24 million speakers, 20 million of whom are in Romania and 3 million in Moldova. The language can also be heard in Ukraine (400,000 speakers), Macedonia, Voivodina, Albania, and Hungary. The largest Diaspora outside Europe is in the United States of America with over 200,000 speakers.

Conclusion

On the whole, 780 million people speak a Romance language, which represent about 20% of the Indo-European language speakers and 8% of humanity, as shown below.

Western Romania	Italo-Roman	Italian	66,000,000
		Sardinian	1,200,000
	Ibero-Roman	Spanish	300,000,000
		Portuguese	170,000,000
		Galician	2,500,000
	Gallo-Roman	Catalan	6,020,000
		French	200,000,000
		Occitan	10,000,000
	Rhaeto-Roman	Romansch	36,000
		Ladin	30,000
		Friulan	600,000
Eastern Romania	Daco-Roman	Romanian	24,000000
			780,386,000

Speakers of Romance languages are superior in number to that of Germanic languages (about 560 million speakers) thanks, in particular, to Latin America. Spanish is nowadays the most spoken language, followed by French and Portuguese. Still, they cannot quantitatively vie with Chinese (at least 900 million) or English, which, although officially spoken by only 450 million people, has become the international language for communication and information.

References

Allières, J. 2000. Les langues de l'Europe. Paris: PUF, Que sais-je ?.

Bach, K.F., and G. Price. 1977. Romance linguistics and the Romance languages: a bibliography of bibliographies. London: Grant and Cutler.

Bal, W. (ed.) 1992. Les langues régionales romanes de Wallonie. Bruxelles, Traditions.

Baldelli, I. (ed.) 1987. La lingua italiana nel mondo. Rome: Istituto della Enciclopedia italiana.

Bec, P. 1980. La langue occitane. Paris: PUF, Que sais-je ?.

Boléo, M. de P. 1974–1975. Estudos de linguística portuguesa e românica, 2 vols. Coimbra: Acta Univ. Conimbrigensis.

Chaurand, J. 1972. Introduction à la dialectologie française. Paris: Bordas.

Entwisle, W.J. 1962. The Spanish language (together with Portuguese, Catalan and Basque). London: Faber & Faber.

Ferreira, J. de A. 1989. Bibliografia selectiva da língua portuguesa. Lisboa: Instituto de Cultura e Língua Portuguesa.

Foster, D.W., and V. R. Foster. 1970. Manual of Hispanic bibliography. Seattle: University of Washington Press.

Griera, A. 1947. Bibliografía lingüística catalana. Barcelona: Escuela de filología.

Hall R.A. Jr. 1958–1988. Bibliografia della linguistica italiana, 6 vols. Firenze: Sansoni.

Harris, M., and N. Vincent. 1977. The Romance languages. London: Routledge.

Lleal, C. 1990. La formación de las lenguas románicas peninsulares. Barcelona: Barcanova.

Lombard, A. 1974. La langue roumaine. Paris: Klincksieck.

Martin, R., and M. Wilmet. 1973. Guide bibliographique de linguistique française. Paris: Klincksieck.

Maxfield, M.F. 1941. Raeto-Romance bibliography. Chapel Hill: University of North Carolina Press.

Melo, G.C. de –. 1981. A lingua do Brasil. Rio de Jeneiro: Padrão.

Pei, M. 1941. The Italian language. New York: Columbia University Press.

Picoche, J., and Chr. Marchello-Nizia. 1989. Histoire de la langue française. Paris: Nathan.

Posner, R. 1996. The Romance languages. Cambridge: Cambridge Univeristy Press.

Pottier, B. 1969. Grammaire de l'espagnol. Paris: PUF, Que sais-je ?.

Roca-Pons, J. 1971. Introducció a l'estudi de la llengua catalana. Barcelona: Vergara.

Rohlfs, G. 1977. Le Gascon. Pau-Tübingen: Marrimpour Niemeyer.

Tétu, M. 1987. La Francophonie. Histoire. Problématique. Perspectives. Montréal: Guérin.

Vasconcellos, J.L. de –.1970. Esquisse d'une dialectologie portugaise. Lisboa: Centro de Estudos filológicos.

Veny, J. 1982. Els parlars catalans : síntesi de dialectologia. Palma de M: Moll.

Wagner, M.L. 1951. La lingua sarda : storia, spirito e forma. Bern: Francke.

Walter, H. 1994. L'aventure des langues en occident. Paris: Laffont.

Wartburg, W. von –. 1967. La fragmentation linguistique de la Romania. Paris: Klincksieck.

Ethnology Database: www.ala.doc.ic.ac.uk/~rap/Ethnologue/wgt.html

Titus: www.fusl.ac.be/Files/General/BCS/Ling1.html

THIERRY PONCHON

See also **Canada; Dialectology; France; French Language; Indo-European 2: Germanic Languages; Indo-European 3: Indo-Iranian Languages; Indo-European 5: Slavic; Italian; Italy; Latin; Meillet, Antoine; Mexico; Phonology; Romanian; Saussure, Ferdinand de; Spain; Spanish and Iberoromance Languages**

Indo-European 5: Slavic

The Slavic languages, spoken by some 288 million people, constitute one of the branches of the Indo-European family. Until approximately 500 CE, before it disintegrated into the various Slavic languages spoken today, the language of the Slavs was relatively uniform. This language, referred to as Proto-Slavic or Common Slavic, was not written down; thus, it must be reconstructed by using the evidence of later written and oral sources. The phase of the unattested Slavic language up to 500 CE is usually referred to as Proto-Slavic, and from that point to approximately the tenth century, it is referred to as Common Slavic. Its closest relative is the Baltic family, made up of modern Latvian and Lithuanian, as well as extinct Old Prussian. The modern Slavic languages are divided into three branches: West (Czech, Slovak, Upper and Lower Sorbian, Polish, and Kashubian), East (Russian, Ukrainian, and Belarusian), and South (Slovene, Croatian, Serbian, Bosnian, Macedonian, and Bulgarian). Today, the Slavic languages are spoken over a large part of Europe and parts of northern and central Asia, as well as in diaspora communities in North and South America and Australia. Most are national standard languages: Polish (Poland), Czech (Czech Republic), Slovak (Slovakia), Russian (Russia, also as a second language in many former republics of the erstwhile Soviet Union), Belarusian (Belarus), Ukrainian (Ukraine), Slovene (Slovenia), Croatian (Croatia), Bosnian (Bosnia and Herzegovina), Serbian (Serbia and Montenegro), Macedonian (Macedonia), and Bulgarian (Bulgaria). Upper and Lower Sorbian are spoken in Germany by the ethnic Sorbian minority in and around the towns of Bautzen and Cottbus, respectively. Kashubian, a divergent dialect of Polish that has gained literary status, is spoken in northern Poland in and around the city of Gdansk. Significant minority populations of Slavs exist outside the borders of the matrix nations, e.g. Russian is spoken by significant percentages of the population of Belarus, Estonia, Latvia, and Ukraine; Slovak is spoken in Hungary and Ukraine; and Slovene is spoken in Italy, Austria, and Hungary.

Slavic writing began in the late ninth century CE in connection with the conversion of the pagan Slavs to Christianity and, specifically, the mission of the Byzantine monks, Constantine and Methodius, who developed the first alphabet for the Slavs, called Glagolitic, and translated the Scriptures into Slavic. The earliest surviving texts are from the late tenth century; this means that there are no extant texts produced during the lives of Constantine and Methodius. By the eleventh century, the Glagolitic alphabet was largely replaced by a modified version of the Greek alphabet, which has come to be known as Cyrillic. This alphabet is still used in modernized forms by the East Slavs, as well as the Serbs, Macedonians, and Bulgarians. The

bulk of the earliest texts that have survived are from the eleventh century, the language of which is now referred to as Old Church Slavic (or Old Church Slavonic, which is also the British term), because of their religious content and function. However, their writing was also used for secular purposes, as evidenced by the Novgorod Birchbark Letters, which contain business records and personal correspondence. The Letters are still being excavated in and around the town of Novgorod, Russia. Latin script writing developed among the Slavs who used the Roman rite, and it continues to be used in modified, modern forms by the Czechs, Slovaks, Poles, Kashubians, Slovenes, Croats, and Bosnians. The oldest use of the Latin alphabet for a Slavic language is attested by the Freising Folia—three brief texts in the Roman rite that reflect an early eleventh-century Slavic dialect that was a precursor to modern Slovene.

The Slavic languages are most closely related to the Baltic languages, as evidenced by certain innovations that are common to the two families and absent in other Indo-European branches. For example, both Slavic and Baltic have abstract nouns formed with the component *-iba*: Old Russian *družьba* 'friendship', Latvian *draudzība* 'friendship' (the Indo-European root **dhreugh-* 'to keep together' was derived as *drug* 'friend' in Russian and other Slavic languages; it is also related to Old English *gedrēag* 'pack [of animals]'). There is some debate as to whether these common innovations arose as a consequence of a continuous common dialect emerging from Indo-European or a rapprochement resulting in intensive contact between Proto-Slavic and Proto-Baltic. Whether through common origin or contact, the close relationship of Slavic and Baltic continued up to about the fifth century BCE, a time depth at which it is very difficult to reconstruct the early forms of the language in adequate detail. Before the Balto-Slavic period, the precursor of Slavic seems to have emerged from Indo-European in conjunction with the precursors of Indo-Iranian, Armenian, Germanic, and Celtic. After the Balto-Slavic period, having migrated westward from the Indo-European homeland, Proto-Slavic crystallized as a separate ethno-linguistic entity in what is now Ukraine, in an area bounded by the rivers Bug in the west, the Dnepr in the east, and the Pripet' in the north (separating them from the Proto-Balts), although the exact location of the Proto-Slavic homeland cannot be determined.

Borrowings of words from non-Slavic languages that are attested throughout the Slavic world give evidence of the contacts that the Proto-Slavs had before their migrations. Among the earliest of these contacts were with Indo-Iranian-speaking peoples, the Alans, Scythians, and Sarmations, with whom the Proto-

Slavs apparently shared religious beliefs: *bogŭ* 'god', *rajĭ* 'heaven', *svętŭ* 'holy', *xvala* 'glory'. Germanic loanwords entered Slavic over a longer period, before, presumably during, and after the migrations. Early loans from Proto-Germanic and Gothic include *kupiti* 'to buy', *t'ud'ĭ* 'foreign', and *stĭklo* 'glass'. Later loans from Old High German, but still in the Common Slavic phase, include *kral'ĭ* 'king' (from *Karl*, referring to Charlemagne) and *penędzĭ* 'coin' (cf. German *Pfennig*). Borrowings from various Turkic languages appear throughout Slavic, although they are arguably from different sources, e.g. Russian *kolpak* 'hat' vs. Czech *klobouk*, Slovene *klobuk* (cf. Crimean Tatar *kalpak* 'cap'). Greek and Romance loanwords entered Slavic substantially through the mediation of the Church, more or less at the beginning of the historical period (i.e. from the tenth century), e.g., *dĭjavolŭ* 'devil', *psalmŭ* 'psalm', *krĭstŭ* 'cross' (Greek); *mĭša* 'Mass', *križĭ* 'cross', *židŭ* 'Jew' (Latin). Contact with varieties of Romance and other languages is responsible for many of the peculiarities of the Balkan languages. Contact with Baltic Finnic languages played a role in the formation of Russian, especially its northern dialects.

The Slavic languages present a variegated picture with respect to their sound structures. The West and East Slavic languages are characterized by complex distinctions in consonants. Russian, for example, distinguishes between plain and palatalized consonants, the latter of which are characterized by a raised tongue position, similar to that produced by English speakers when pronouncing the letter *y,* viz. *privedëš'* [prʸivʸi'dʸoʃ] 'you will bring'. However, it should be emphasized that Russians perceive this secondary articulation as a characteristic of the consonant itself, because they contrast such sounds with sequences of plain consonant plus *y,* e.g. *sel* ['sʸel] 'sat' and *s"el* ['sʸyel] 'ate'. Czech has intensified the development of palatalization so that what were formerly palatalized consonants have evolved into consonants articulated at the middle of the palate; thus, the first consonant in the word *tichý* 'quiet' is neither *t* nor *k*, but a sound in between the two; this sound contrasts with a plain *t*, as in *ty* 'you'. This intensification has resulted in the development of a trilled fricative, *ř* (*řeka* 'river', cf. Russian [rʸe'ka]), which is pronounced as the *s* in English *plea<u>s</u>ure* and simultaneously rolled/trilled as in Spanish *bu<u>rr</u>o*. In Polish, this sound existed historically, but it has merged with the nontrilled fricative, *rzeka* ['žeka] 'river'. Vocalic and accentual systems vary considerably from language to language. Czech and Slovak distinguish long and short vowels and have stress fixed on the first syllable of the word, e.g. Czech *dal* ['dal] 'he gave', *dál* ['da:l] 'further'. Slovene and Serbo-Croatian also preserve long and short vowels,

but in addition they distinguish between rising and falling intonation (pitch) in stressed syllables, e.g. Slovene *brati* [b'rá:t] 'to read' (long low or rising pitch), *brat* [b'rà:t] 'to go read' (long high or falling pitch), *bràt* [b'ràt] 'brother' (short high or falling pitch). In general, there seems to be a complementary relationship between consonantal and vocalic/accentual systems: the more complex the consonantal system, the simpler the vocalic and accentual systems, and vice versa.

As regards word structure, the Slavic languages are characterized by inflection; that is, words change to reflect grammatical relationships. For example, the Russian word *sobaka* 'dog (as a subject)' has the following forms: *sobaki* 'of the dog', *sobake* 'to the dog', *sobaku* 'dog (as an object)', *sobakoj* 'as a dog', *na sobake* 'on the dog', *sobaki* 'dogs (subject)', *sobak* 'of the dogs' or 'dogs (as an object)', *sobakam* 'to the dogs', *sobakami* 'as dogs', and *na sobakax* 'on the dogs'. This characteristic allows word order to remain flexible, which in turn permits word order to take on functions other than indicating grammatical relations, such as, for example, emphasis: *Ja ljublju sobaku* 'I love the dog' (normal word order) vs. *Sobaku ljublju ja* 'It is I who loves the dog' (emphasizes the subject). In the Balkan Slavic languages, Bulgarian, and Macedonian, most of these inflectional changes in nouns have been eliminated. Slavic verb forms mark person (the subject of the verb) and number (singular, plural, and in some Slavic languages, dual); for this reason, many Slavic languages omit the subject pronoun in neutral speech, e.g. Czech *vidím* 'I see', *vidíš* 'you see', *vidí* 's/he sees', *vidíme* 'we see', *vidíte* 'you see' (polite or plural), *vidí* 'they see'. Three genders are distinguished—masculine, feminine, and neuter—although unlike many Western European languages, these are not expressed by definite articles ('the', such as German *die, der, das* or French *la, le*), there being no articles in Slavic languages. Rather, noun endings indicate gender, e.g. Russian *žena* 'wife', *kniga* 'book' (both feminine); *čelovek* 'person', *stol* 'table' (no ending—both masculine); and *okno* 'window' (neuter), *more* 'sea'. Gender serves to coordinate sentence elements through agreement; for example, an adjective modifying a noun must agree in gender (as well as other grammatical categories) with the noun: Russian *krasivaja žena* 'beautiful wife' (feminine), *krasivoe more* 'beautiful sea' (neuter). It may also agree with the past tense of a verb, e.g. *žena stojala tam* 'the wife was standing there', *čelovek stojal tam* 'a person was standing there'. Moreover, the formation of gendered noun pairs is effected by adding a suffix to produce the feminine member of the relationship, e.g. Russian *kurd* 'Kurdish man', *kurdjanka* 'Kurdish woman'. A typical characteristic of Slavic inflection and word formation is the rich array of alternations in the shape of words; these alternations have led to a well-developed investigation into the nexus of sound systems and word structure known as morphophonemics. An example of a morphophonemic alternation is found in Russian *vižu* 'I see' vs. *vidiš* 'you see', where the basic form of the root *vid-* 'see' changes to *viž-* in the context of the first-person singular nonpast.

Verbs in Slavic languages are distinctive among European languages in that they overtly mark aspectual contrasts. That is, they distinguish completed and uncompleted actions/events, as well as express other varieties of temporal and spatial manners of organization. For example, Slovene *skočim* 'I jump (once)' (completed—perfective aspect), *skačem* 'I am jumping' or 'I jump habitually' (e.g. as a professional ski-jumper) (uncompleted—imperfective aspect). Further aspectual distinctions may be derived through suffixation and prefixation, e.g. Russian *ona brosila mjač* 'she threw the ball' (perfective), *ona brosala mjač* 'she was throwing the ball' (imperfective); *ona podbrosila mjač* 'she tossed the ball up' (perfective), and *ona podbrasyvala mjač* 'she was tossing the ball up' The arrangement of tense (time) in Slavic verbs has been considerably reorganized from the Indo-European starting point. Russian is an example of one of the more innovative tense systems. For example, the future is formed by using a perfective verb with the same formal properties as the imperfective present: *ja pojdu* 'I shall go' (the corresponding imperfective *ja idu* means 'I am going' or 'I go'). The future with the imperfective aspect is formed by adding an auxiliary verb *budu* (which goes back to an Indo-European root **bheuH-* meaning 'to grow') to the infinitive: *ja budu iti* 'I will be going.' Past tense is formed by a participle derived with the formant *-l* that agrees in gender and number with its subject, e.g. *ja pošël* 'I left.' This rather simple schema supplanted a much more complex system of past tenses that was still in place in medieval varieties of Slavic, e.g. Old Russian (twelfth to thirteenth centuries), *i vŭlny byša vyše korablä* 'and the waves were higher than the ship' (aorist, a simple past narrated event); *i reče sĭ unoša roda velika jestĭ bylŭ* 'and this youth said that he was of a great family' (perfect tense, meaning that the narrated past event is of relevance to the moment of narration). This complex system was preserved and developed further in Macedonian and Bulgarian, which have otherwise lost much of the complexity of noun and adjective inflection found elsewhere in Slavic. Parallel to the complementariness between consonants and vowels/accent, pointed out previously, such a relationship seems to hold in Slavic languages between the complexity of nominal and verbal inflection.

References

Birnbaum, Henrik. 1979. Common Slavic. Progress and problems in its reconstruction. Columbus, OH: Slavica. xii+436pp.

Birnbaum, Henrik, and Peter T. Merrill. 1985. Recent advances in the reconstruction of common Slavic (1971–1982). Columbus, OH: Slavica. vi+141pp.

Comrie, Bernard, and Greville G. Corbett (eds.) 1993. The Slavonic languages. London: Routledge. xiii+1078pp.

Gołąb, Zbigniew. 1992. The origins of the Slavs. A linguist's view. Columbus, OH: Slavica. 454pp.

Schenker, Alexander M. 1995. The dawn of Slavic. An introduction to Slavic philology. New Haven: Yale University Press. xviii+346pp.

Schenker, Alexander M., and Edward Stankiewicz (eds.) 1980. The Slavic literary languages: formation and development. New Haven: Yale Concilium on International and Area Studies. x+287pp.

Shevelov, George Y. 1965. A prehistory of Slavic. The historical phonology of common Slavic. New York: Columbia University Press. xx+62pp.

MARC L. GREENBERG

See also **Balkans; Baltic Languages; Polish and West Slavic Languages; Russian and East Slavic Languages; Serbo-Croatian and South Slavic Languages; Soviet Union; Soviet Union: Successor States**

Indo-Pakistani Sign Language

Indo-Pakistani Sign Language (IPSL) is a visual-gestural language that uses movements of the hands, facial expressions, and head/body positions to convey linguistic messages. Dialects of IPSL are used in deaf communities in urban centers of the Indian Subcontinent. Figure 1 shows the extent of the geographic area as documented to date. It is likely that dialects of IPSL are also used in other parts of India and/or Pakistan, maybe even in neighboring countries (Sri Lanka, Nepal, Bangladesh), but this has not been fully documented.

In the deaf community, the sign language is simply called 'sign/signing/sign language', sometimes in combination with the sign for the country (as in *INDIA SIGN*). Various names are used by hearing people. In the Hindi/Urdu-speaking area, the sign language is known as *ishaaron kii zubaan* ("language of signs"). Official usage also refers to "Pakistan Sign Language (PSL)" and "Indian Sign Language (ISL)", respectively, although this usage runs contrary to the linguistic facts.

The large area covered by one and the same sign language is particularly noticeable in view of the great linguistic diversity of spoken languages in the region. IPSL is used in both the Indo-European language area (e.g. Hindi-, Nepali-, and Marathi-speaking areas) and the Dravidian language area (e.g. Telugu-speaking areas in Andhra Pradesh). All IPSL dialects have the same grammar, but lexical variation may be considerable. On average, IPSL dialects have about 75% of shared vocabulary, with about 25% of the vocabulary different across dialects.

Despite dialectal differences, IPSL users can communicate freely across a large geographic area. Many are multidialectal or become multidialectal very quickly as soon as they travel to other areas. Moreover, since IPSL is a minority language in constant contact with spoken languages, all IPSL users are to some extent bilingual in the signed and the spoken medium, the latter mostly in its written form. However, competence in spoken languages, be it English, Hindi, or one of the regional languages, is often very low among deaf people. The use of so-called 'mouthing' is also a contact phenomenon resulting from the bilingual situation. Mouthing means that sign language users accompany signs with mouth movements that correspond to words of a spoken language. For example, one may imitate the mouth movements of the Hindi word *kaam* 'work' while signing *WORK*.

The size of the language community has not been documented reliably, but IPSL users definitely number in the hundreds of thousands, possibly even over a million, thus representing one of the largest sign language communities in the world. IPSL is not known to be genetically related to any other sign language. A minor influence from British Sign Language can be seen, for example, in the use of a two-handed manual alphabet (fingerspelling) for representing English words. No manual alphabet for indigenous Indian languages is widely in use.

Word Classes and Sentence Structure

IPSL has three main word classes: verbs, multifunctional words, and particles. There are no word classes of nouns, adjectives, and adverbs. The most important

Figure 1. Geographic extent of IPSL.

subclass among the particles is a set of nine functional particles. These appear at the end of a sentence or clause and indicate sentence types such as negative sentences, questions, completed actions, commands, and existentials (equivalent to English 'there is').

The class of verbs is rather small, consisting of about 50 signs, such as the signs for 'help', 'ask', 'give', 'tell', 'teach', 'go', 'come', and the like. These are characterized by so-called 'directional movement', which means that the movement of the verb varies depending on the relationship between the persons, objects, or places involved. Figure 2 shows an example of a sign expressing a relationship between two persons.

Most signs in IPSL belong to the class of multi-functional signs, including signs as diverse as 'change', 'begin', 'difficult', 'black', 'three', 'self', 'many', 'war', 'child', and so on. One and the same multifunctional sign can function as the equivalent of both a noun and a verb/predicate in other languages. For example, one would sign *BEGIN DIFFICULT* to express 'The beginning is difficult.' (nominal function), and *WAR BEGIN COMPLETIVE-PARTICLE* to express 'The war has begun.' (verbal function).

IPSL has a subclass of signs related to pointing. Pointing with the index finger is often functionally equivalent to pronouns in spoken languages. Pronouns have the full set of IPSL number distinctions, which include a form unspecified for number that may have either singular or plural reference, a dual (for the number 'two'), and several types of plurals. Another type of pointing sign, with the index finger drawing a line between two locations in space, is used as an auxiliary. The auxiliary can express subject–object-like relationships in a general way (translatable as 'he to me', 'I to you', and so on) in combination with signs that do not have directional movement, for example: *SIGN AUXILIARY-you-me-you* 'You talk to me in sign language and I talk to you' (see Figure 3).

IPSL is a predicate-final language, the basic sentence structure being ARGUMENT – PREDICATE – FUNCTIONAL PARTICLE, as in *SCHOOL FAR NEGATIVE-PARTICLE* 'The school is not far.' A second argument will usually be a pronoun and can appear just about anywhere, e.g. *INDIA YOU LIKE EXISTEN-TIAL-PARTICLE*? 'Do you like India?' Sentences are often highly elliptic, leaving out information that can
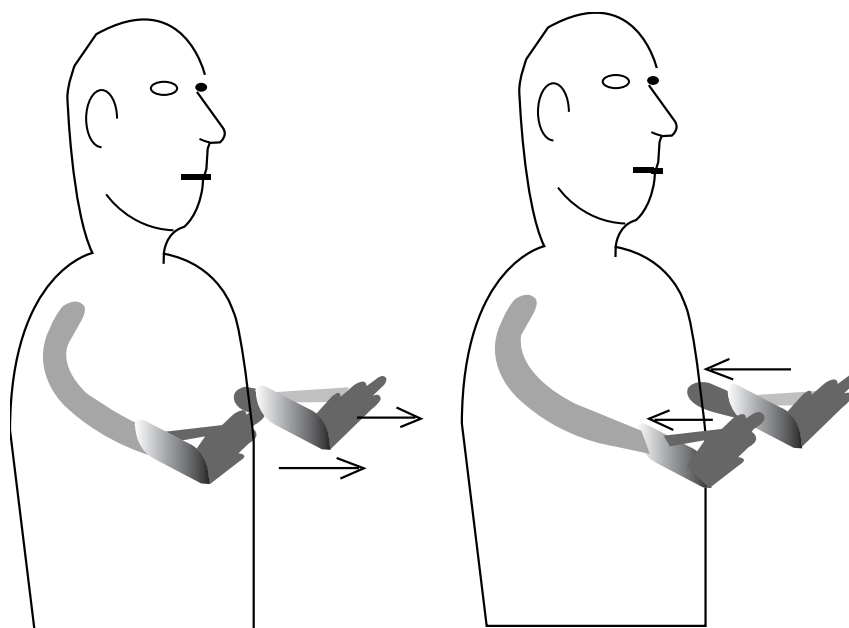
Figure 2. 'I give someone a gift' and 'Someone gives me a gift'.
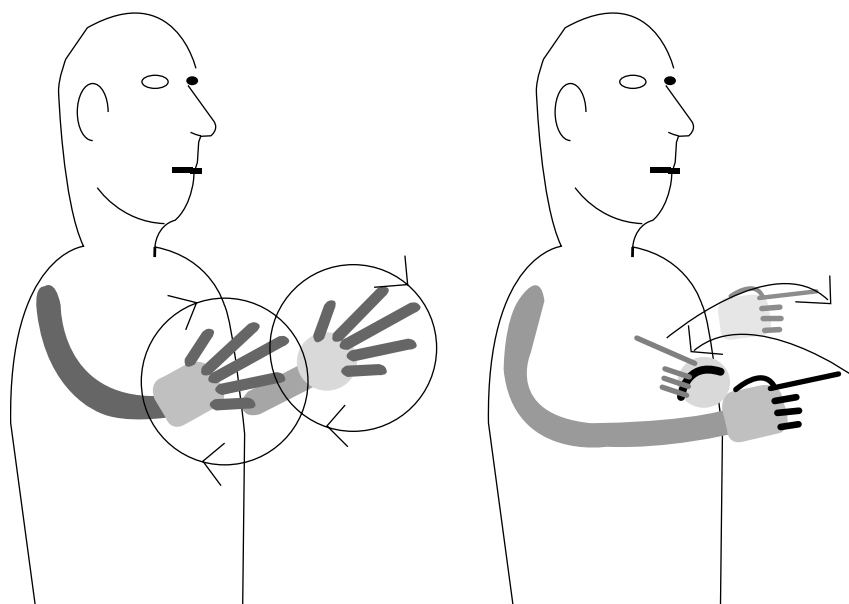
Figure 3. 'You talk to me in sign language and I talk to you'.

be recovered from the context, so that the predicate is the only obligatory constituent. Signs indicating time (such as 'tomorrow', 'before', 'in the future', etc.) always appear at the beginning of a sentence. Like most other sign languages, IPSL has no tense system. Rather, time is expressed at the level of the text, with time signs placed at the beginning of a new paragraph each time the time reference changes.

Structural Features Shared with Other Sign Languages

IPSL shares a number of aspects of its structure with other sign languages. By modifying the movement pattern of a sign, it is possible to express several aspectual meanings, such as 'do something repeatedly', 'be about to do something', or 'something happening
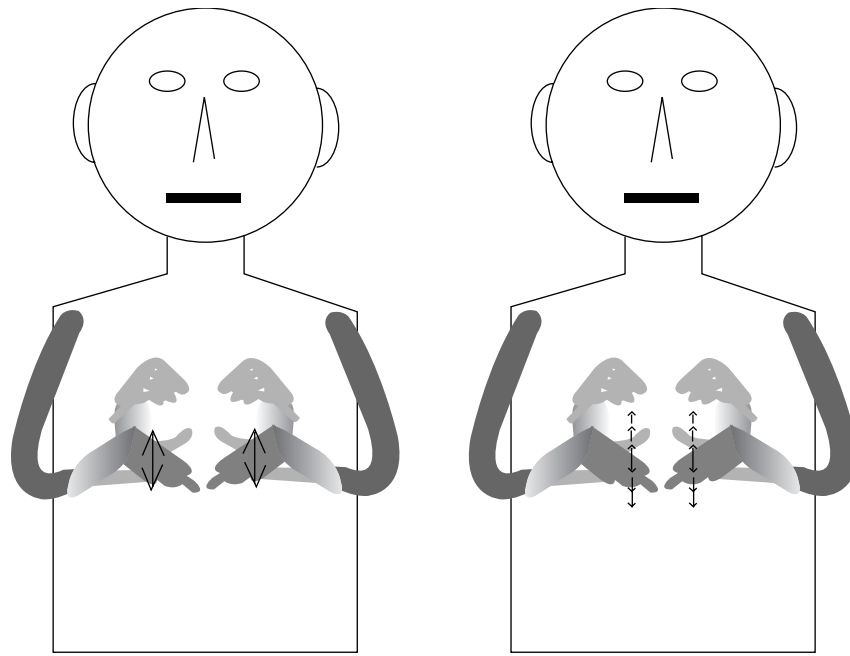
Figure 4. 'Develop': single opening movement; and 'develop gradually': stepwise gradual opening.

gradually'. Figure 4 shows an example, and it is obvious how the form of the sign corresponds to its meaning. Such iconicity, both in the vocabulary and in grammatical processes, is also a characteristic feature of sign languages in general and is used extensively in IPSL.

Iconicity also plays a role in some of the most productive IPSL constructions. These include a large set of signs denoting geometrical shapes, such as 'square two-dimensional', 'round three-dimensional', 'flat surface', and so on. They can be used in countless modifications and combinations to describe a wide variety of objects in terms of their visual-geometric properties. IPSL also has a subsystem known as 'classifiers' in other sign languages. However, the IPSL classifiers are fewer in number, used less frequently, and refer to people and animals only, not to other categories such as vehicles. In such a construction, one sign can express the equivalent of a whole sentence in English (see Figure 5).

IPSL and other sign languages make use of a wide range of facial expressions for both adverbial concepts (such as 'do with difficulty', 'more and more', 'unfortunately', etc.) and grammatical categories. In IPSL, the latter include facial expressions that go with yes/no-questions, questions with question words (e.g. 'what', 'how'), negative sentences, and subordinate clauses. For example, in questions with question words, the eyebrows are raised and the head is tilted backward. Such facial expressions, which are often obligatory and have to follow specific syntactic rules,
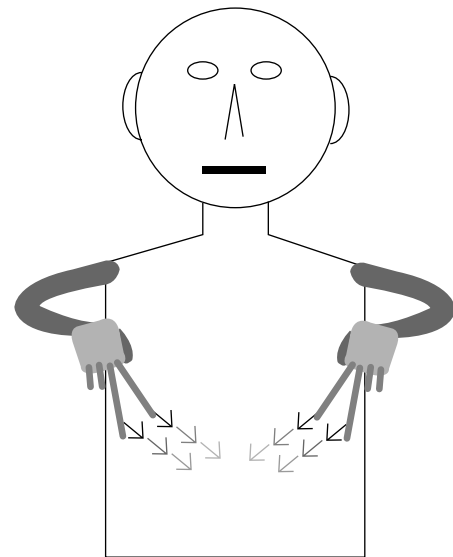
Figure 5. 'Two persons approaching each other'.

are extremely important in all sign languages documented so far.

Finally, the important grammatical role of space, for example, with directional movement, is a unique feature of sign languages that has no correlate in spoken languages. In a text, sign language users often assign particular points in space to the objects or people they are talking about and make continuous reference to these locations, for example, by pointing at them. This process is known as 'localization' and

plays an important role in IPSL grammar as well as in other sign languages.

Structural Features Different from Other Sign Languages

On the other hand, many structural features are peculiar to IPSL and differ from other sign languages. For example, many known sign languages do not have an auxiliary construction of the type described above for IPSL. In the negative, IPSL has two different functional particles for neutral negation (saying that something is not the case) and contrastive negation (rejecting something that has been said or implied before). Unlike in other sign languages, there is no particular construction for conditionals ('if-clauses'), and there are no conjunctions to link clauses. Instead, there is a general subordinating construction that is marked nonmanually by a combination of facial expression, head posture, and sign rhythm.

IPSL has few genuine compounds, but there are a number of striking types of sign combinations. Almost all signs for kinship terms are compositional, with the sign for 'man/male' or 'woman/female' preceding a sign for the family relationship, for example, *MAN+SIBLING* 'brother', *WOMAN+MARRY* 'wife' (see Figure 6). IPSL has a single question word with a general interrogative meaning (see Figure 7). To ask specific questions, the general interrogative must be combined with other signs, for example, *FACE+QUESTION-WORD* 'who', *PLACE+QUES-TION-WORD* 'where', *TIME+QUESTION-WORD* 'when'.

The use of mouthing in IPSL is also markedly different from other sign languages. Since there are so many different spoken languages in the IPSL-using area, one and the same sign may be accompanied by mouthing from different spoken languages in different areas. While in some sign languages, mouthing is functionally very important, it is much less significant in IPSL. The extent of mouthing is extremely variable; hence, signing with extensive mouthing and signing with no mouthing at all is equally possible.

Even the signing space in IPSL is structured differently from other sign languages. The upper sign space around the head and above the shoulders is particularly significant and is associated with the concepts of distance and authority. Therefore, when place names or names of institutions with authority are mentioned, an index finger pointing is added that has to be directed upward, for example, *GOVERNMENT INDEX-up, CALCUTTA INDEX-up.*

Social and Political Factors

The deaf community in India and Pakistan is a linguistic and cultural community rather than an ethnic or political entity. Focal points are the deaf schools and the deaf associations, of which there are many in all urban centers. On the other hand, the rural areas are usually cut off from the deaf community due to a lack of infrastructure, and isolated deaf people in these areas use limited ad hoc gestural communication ('home sign'). The origins of the deaf educational system date back to the late nineteenth century, when the first schools for the deaf were founded in India (c. 1882 in Mumbai, c. 1893 in Calcutta). However, this does not necessarily coincide with the age of the sign language itself. So far, there is no reliable information about when and how IPSL originated.

IPSL is not an officially recognized language in any part of the Indian Subcontinent, and there is a great deal of prejudice and misinformation about it. This
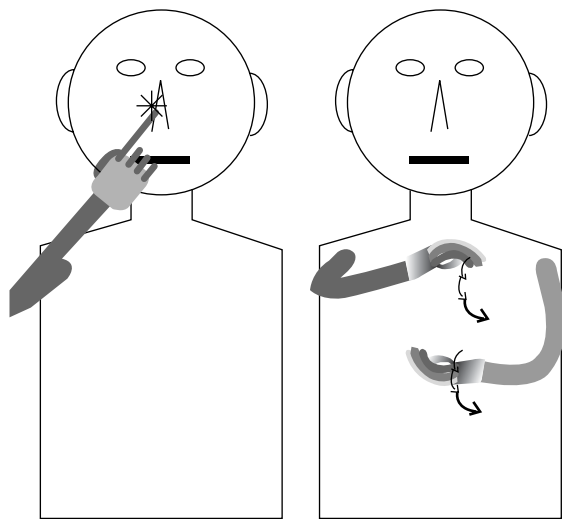
Figure 6. 'Daughter': WOMAN+BORN.
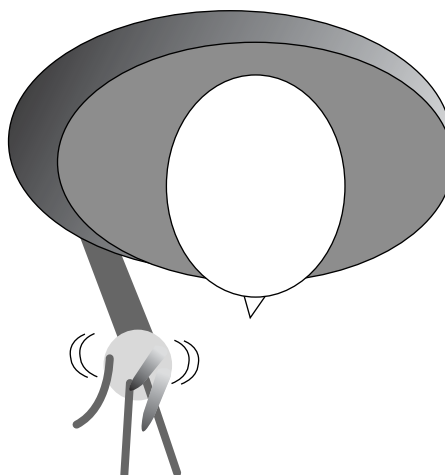
Figure 7. General question word.

may account for various recent attempts at 'developing' the sign language, for example, by inventing contrived sign systems ('Indian Sign System', 'Sign Urdu', etc.). In these systems, which are not used by deaf people among themselves, signed sentences are constructed to mirror the structure of a spoken language, disregarding IPSL structures and adding a number of newly invented signs. There have also been equally unsuccessful attempts at 'standardizing' IPSL dialects.

Deaf people themselves have often adopted the negative views on sign language held by the hearing majority. Use of the primary sign language, in particular, in the deaf educational system, is still stigmatized, although the situation is improving. Officially, most schools for the deaf follow the policy of 'oralism', with emphasis on speech training/articulation and lip reading, but day-to-day practice is often more open toward forms of manual communication. Over the last few years, the Indian government has supported the development of sign language courses and the training of qualified sign language teachers, so that attitudes toward sign language and related policies are beginning to change. However, the struggle for a genuine recognition of the linguistic human rights of the deaf community in the Indian Subcontinent still continues.

References

Ayjnihh and Ulrike Zeshan. 2003. Advanced course in Indian sign language. Mumbai: Ali Yavar Jung National Institute for the Hearing Handicapped.

Jepson, Jill. 1991. Two sign languages in a single village in India. Sign Language Studies 20(70). 47–59.

Zeshan, Ulrike. 2000. Sign language in Indo-Pakistan: a description of a signed language. Amsterdam: Benjamins.

———. 2000. Gebärdensprachen des indischen Subkontinents (Sign languages of the Indian subcontinent). Munich: LINCOM.

———. 2001. Mouthing in Indopakistani sign language: regularities and variation. The hand is the head of the mouth: the mouth as articulator in sign languages, ed. by Penny Boyes Braem and Rachel Sutton-Spence. Hamburg: Signum.

———. 2003. Indo-Pakistani sign language grammar: a typological outline. Sign Language Studies 3(2). 157–212.

An annotated historical bibliography of deafness and sign language in the region, compiled by Mike Miles, is available at <http://www.sign-lang.uni-hamburg.de/bibweb>.

ULRIKE ZESHAN

See also **American Sign language; British Sign language; Japanese Sign language; Signed Languages**

Inflection and Derivation

Inflection and derivation are terms used in morphology, the study of word structure. They distinguish the two main ways of forming words in fusional languages such as Finnish, Italian, Arabic, and—to a lesser extent—English. Isolated languages such as Mandarin, Thai, and Vietnamese tend not to use affixes; i.e. these terms do not apply very well to these languages. An affix is a morpheme attached to a word to change its grammatical function or meaning: at the start (prefix), around it (circumfix), in the middle (infix), or at the end (suffix).

Inflectional affixes signal grammatical relationships such as number, person, tense, and case, and they do not change the grammatical class of the word to which they are attached. In highly inflected languages such as Arabic, much information is conveyed very economically, as shown by the paradigm of the perfect form of *yaktub*:

katabtu	I have written
katabti	you (feminine singular) have written
katabta (masculine singular)	you have written
katabat	she has written
kataba	he has written
katabna	we have written
katabtunna	you (feminine plural) have written
katabtum	you (masculine plural) have written
katabnaa	they (feminine plural) have written
katabuu	they (masculine plural) have written
katabtumaa	you two have written
katabataa	they two (feminine) have written
katabaa	they two (masculine) have written

The appearance of inflectional suffixes in a language interacts with sentence formation. The more highly inflected the language, the less important the order of words to convey meaning. In Latin, the word order in the sentence *parvum puerum magna puella vidi* could be changed without loss of grammatical meaning because the inflectional suffixes indicate 'the tall girl saw the small boy'and not vice versa: *-a* is subject case, feminine, third-person singular, whereas *-um* is object case, masculine, third-person singular.

529

Although Old English inflected all nouns, pronouns, and adjectives, there are only eight inflectional suffixes in Modern English, only one of which can be attached to a word at a time:

-s	Third-person singular present (*wants*)
-s	Plural noun (*dogs*)
-ed	Past tense (or past participle) (*wanted*)
-'s or *-s'*	(*dog's, dogs'*) Genitive case of noun
-en	Past participle (*given*)
-ing	Progressive aspect (*giving*)
-er	Comparative adjective (*bigger*)
-est	Superlative adjective (*biggest*)

These inflections are highly 'productive' —that is, new nouns, verbs, or adjectives that enter the language will be automatically assigned these markings. 'Nonproductive' patterns lack this potential; for example, alternate plural markings such as *-en* (oxen) or *-ren* (children) would not be assigned to new nouns.

However, the limited number of inflectional affixes in English means that they are overworked and potentially ambiguous. For example, the suffix *-s* on *trains* can mark the third person of the verb *train* or the plural of the noun *train*. Similarly, *train's* can indicate possession, as in *the train's noise,* or it can be a contracted form of a noun and auxiliary, as in *the train's here.* Furthermore, *-ed*, *-ing*, and *-en* may be derivational adjectival markers (*interested*, *interesting*, *broken*), and *-er* may mark an agent of a verb (*worker*, *teacher*).

Derivational affixes are used to create new words and, unlike inflectional suffixes, they change the grammatical class of the root word to which they are attached. In most languages, derivational affixes are usually found nearest to the root, whereas inflectional affixes tend to be on the periphery of a word, usually at the beginning or at the end.

In English, most derivational affixes are suffixes, and—unlike inflectional suffixes—more than one can be added to a root. Thus, the noun *nation* combines with *-al* and *-ly* to produce an adjective and then an adverb, respectively. Derivational suffixes can also change subclasses of words: thus, the abstract noun *fatherhood* derives from the concrete noun *father*.

The rules of derivational word formation are complex, but every adult native speaker of a language knows and uses them to understand words not previously encountered. Moreover, many new words can be coined by derivational affixes. For example, a British Prime Minister gave rise to thatcherism and thatcherite, and thatcherize and thatcherizable would be fairly readily understood: these two suffixes (*-ize* and *-able*) are particularly productive. Others (for example, the prefix *im-*, as in *impossible*) are referred to as 'semiproductive' because they can be attached to fewer words. In most languages, derivational word formation is less productive than inflectional. Since the Middle Ages, most changes in English word structure have occurred by attaching derivational suffixes to Latinate roots. However, a recent tendency toward Anglicization may be noted: thus, contemporary additions to the lexicon include compound verbs such as *downgrade*, *downsize*, *outpace*, *outsource*, etc., rather than their Latin-based equivalents.

There is an interaction between affixation and sound structure. For example, in English, the pronunciation of *-s* depends on the previous consonant sound: *-s* is pronounced [s] in *cats*, because the preceding *t* is voiceless, but it is pronounced [z] in *kids*, because the preceding *d* is a voiced sound. Sometimes, the word stress shifts, for example, *photograph/photography*. In other cases, the vowel sound may alter, as in *divine/divinity*. In some cases, a feature of the final consonant changes, for example, *advice/advise*; occasionally, there are changes in both vowel and consonant quality, as in *breath/breathe*.

The previous discussion has made two assumptions. The first is that the two categories of affixation are easily distinguishable—but this is not always so. As noted previously, some inflectional affixes may have the same form as some derivational ones (e.g. *-s*: noun plural or verb third-person singular).

Second, it is also assumed that the notion of a word is itself unambiguous—but this is not the case either. The English indefinite article (*a*) and the Kivunjo expression *näikìmlyïïa* ('he is eating it for her') would each be identified by respective native speakers as a single word. Fully understanding what a word is, and therefore how it is formed through affixation, requires knowledge of sound structure, sentence grammar, and the rules of word formation.

References

Bauer, Laurie. 1983. *English word formation*. Cambridge: Cambridge University Press.

Bauer, Laurie. 1992. *Introducing linguistic morphology*. Edinburgh: Edinburgh University Press.

Jensen, John, T.1990. *Morphology: word structure in generative grammar*. Amsterdam and Philadelphia: John Benjamins Publishing.

Katamba, Francis. 1993. *Morphology*. London: Macmillan.

Matthews, Peter, H. 1991. *Morphology*, 2nd edition. Cambridge: Cambridge University Press.

Spencer, Andrew. 1991. *Morphological theory: an introduction to word structure in generative grammar*. Oxford: Blackwell.

Spencer, Andrew, and Arnold M. Zwicky (eds.) 1998. *The handbook of morphology*. Oxford and Malden, MA: Blackwell.

ROGER BARNARD

See also: **Affixation; Word**

Information Retrieval

Information retrieval (IR) is a branch of computer science (Baeza-Yates and Ribeiro-Neto 1999; Salton and McGill 1983; van Rijsbergen 1979; Jackson and Moulinier 2002) that deals with content-based searches for relevant information on very large document collections, usually composed of hundreds of thousands of documents. These documents—the IR term for all kinds of written texts—are contributed by a large variety of authors and cover a wide range of topics, genres, and writing styles, although scientific papers and newspaper articles prevail. The general goal of an IR system, in a narrow sense, is to locate all the documents in the underlying document collection that are relevant with respect to a user query expressing a specific search topic. Given these documents (or bibliographic references to them), the user of an IR system then has to access these documents, read them, and understand their contents in order to solve the problem that led him or her to formulate the specific search topic. More and more, this document-centered view of IR systems is going to be broadened to include the direct location of factual knowledge in documents and procedures for content condensation (summarization) of documents as well. Still, searches are run on unformatted textual material. This marks the borderline to fact retrieval with database systems that perform searches on prestructured, formatted data collections (e.g. relational tables). In particular, the rise of the World Wide Web and the proliferation of machine-readable texts have spurred a strong interest in IR techniques. Because documents in the WWW are often multimedia aggregates, the goal of IR also gradually expands to include the content-based search of audio, video, and graphical data as well, although the primary focus is still on textual data (and the textual description of other media).

A typical IR scenario can be described as follows: A person wants to find relevant documents on guitar players, particularly those playing jazz or rock music. Rather than submitting this search topic to an IR system in colloquial English, it must be translated into a 'formalized' query format. First, the facets of the topic under consideration are designated by search terms: GUITAR, JAZZ, ROCK MUSIC. Second, these terms have to be grouped according to their logical dependencies. The final search query then looks like this: (GUITAR *and* (JAZZ *or* ROCK MUSIC)). This query triggers a search process in the underlying document collection, which can be paraphrased as follows: 'Find me all documents in which the term GUITAR occurs and, at the same time, either the term JAZZ or ROCK MUSIC, or both, occur, too'. The IR program will then return all hits, i.e. documents that exhibit the required co-occurrences of search terms.

This is the classical Boolean approach to IR, in which search terms are combined by Boolean operators, viz. '*and*' or '*or*', sometimes also '*not*'. Such a Boolean query must exactly match with the terms in a document to be retrieved. Despite its merits as the baseline for almost all commercial IR systems, its drawbacks are evident: Boolean logic is a formal language most people are not familiar with and, even worse, Boolean queries for nontrivial search problems can become quite complex; the result set is often very large and usually unordered, all search terms are equally relevant for the query evaluation, and, finally, the search partitions the entire collection into exactly two subsets: documents that satisfy the Boolean query and documents that do not.

There are some disturbing linguistic phenomena that arise through such a surface-oriented approach, particularly concerning lexical variants due to inflection. A very simple approach for dealing with them is the use of the truncation operator '*$*'. Used as a suffix such as in 'SING$', this expression retrieves all documents in which the string 'SING' appears followed by an arbitrary string. Hence, it matches SING, SING*S*, SING*ER*, SING*ING* but also SING*ULAR*, etc. Linguistically more sophisticated alternatives are stemmers, i.e. programs that strip off suffixes based on a small set of rules without the use of dictionaries (thus preventing e.g. the reduction of SING*ULAR* to SING). In this case, the IR system takes over the initiative and the user no longer has to make ad hoc guesses about morphological regularities (as required by the proper placement of truncation operators). However, linguistically more advanced morphological analysis used to cope with the term variation problem is usually not supported by IR systems.

In the basic Boolean approach to IR, all search terms are equally important. This assumption is often inadequate and, hence, retrieval models have been proposed, which incorporate statistical methods for grading the relevance of terms for a document search. There is a very general assumption underlying the use of statistical methodology for IR, which relates the frequency of occurrence of terms to their relevance as content descriptors. This can be expressed more precisely as

follows: determine the absolute frequency of each term j, tf_{ij} (excluding irrelevant, although highly frequent stop words such as 'the', 'for', 'again', etc.) for each document $i \in [1, n]$ in the entire collection (where n gives the total amount of documents in this collection). Then, determine the corresponding relative term frequencies, documentwise and collectionwise. A term is considered relevant if its relative document frequency by far exceeds its relative collection frequency. In order to stress the specificity of a term, we may also add extra weight to those terms that have a high occurrence frequency in few documents and penalize those terms that are widespread and occur rather frequently throughout the entire collection (and, hence, do not discriminate well for searches). In essence, this is achieved by the so-called $tf*idf$ measure (term frequency—inverse document frequency). Its tf part equals tf_{ij} from above, and its idf part is constituted by $\log(\frac{n}{n_i})$, where n yields the size of the document collection and n_i specifies the number of documents in which term i occurs.

There are more advanced methods, e.g. the vector space model, in which a document is represented by an n-dimensional document-term vector, n being the number of different terms (excluding stop words) in the entire document collection. Interestingly, this approach assumes the dimensions of the vector space to be orthogonal, i.e. all terms are treated as being totally independent of each other (which sounds bizarre from a linguistic perspective). The 'similarity' (another more technical notion to express relevance) of a document and a query can then be measured by comparing the corresponding document and query vectors using well-known metrics such as the cosine measure (the higher the cosine value, i.e. the smaller the cosine of the angle of two vectors, the more similar both the items are). Accordingly, we may also determine the similarity between any two documents in the document collection based on a vector comparison between both of them. This can be exploited for the retrieval process in the following way: when a query identifies a relevant document, then another very similar document (which exhibits a high cosine value relative to the already retrieved document) should also be part of the system output. This approach to classifying documents usually requires clustering procedures, which build on the lexical vector representation of documents. In general, IR systems that apply the vector space model generate as output a list of documents, ranked by decreasing relevance in terms of how similar each document is relative to the query, given their vector representations.

A powerful extension to this single-step ranking approach is called *relevance feedback*. Based upon the results of the first retrieval round, the searcher identifies those documents among the top-ranked k documents (where k usually ranges between 20 and 50), which he or she considers relevant (all remaining ones are considered irrelevant, by convention). The terms of the relevant documents are 'added' to the vector representation of the original query, while those from the irrelevant documents are 'subtracted' from it. The entire search is then rerun with the modified query. With two or three iterations, this usually leads to improved search results within the feedback loop.

Statistical models of term association that use vector space representations and various association factors (such as the cosine measure), are a quantitative way of making sense relations explicit, although they do not distinguish between hyponyms, antonyms, or synonyms, etc. A more traditional way of accounting for lexical semantics is the use of thesauri and classification codes in document retrieval. A *thesaurus* is a lexical repository in which terms are linked to other terms via a small set of informal semantic relations, such as synonymous or quasisynonymous terms, broader or narrower terms, related terms, etc. This might be useful for document retrieval in search modes based on *query expansion*: using a broader (i.e. general) term such as INSTRUMENT explicitly linked to a thesaurus will match documents that not only contain INSTRUMENT as a term but also, say, GUITAR, DRUMS, PIANO, or TRUMPET, i.e. narrower (more specific) terms. Hence, it frees users from thinking of *conceptual* variants of search terms (synonyms, more specific or related terms) when they formulate a query.

Classification codes are another vehicle for expressing semantic relations although, usually, only along the general-specific axis. In classifications, concepts are represented by some kind of alphanumerical string (e.g. INSTRUMENT comes as MI0) and more specific concepts are represented using this string as a prefix so that, e.g. GUITAR may appear as MI01 or MI0G. *Text categorization* (*classification*) is then defined as the problem of assigning a suitable set of category codes to each document. Similarly, *indexing* denotes a process by which a document is assigned a set of index terms describing its main topics, with the index terms either being directly extracted from the source document or derived from some controlled vocabulary, e.g. a thesaurus. As far as the retrieval process is concerned, with lists of category codes or index terms being available, these document surrogates are searched rather than the full text of documents.

The above scenario assumes a person actively submitting an ad hoc query to an IR system and awaiting interactive system response. Nevertheless, the same methodologies can also be used for different information supply models. One of these focuses on *document routing* (or, *document filtering*). In this case, a query that is stable over a certain time interval (some

months, or so) is continuously evaluated in the system. When new, incoming documents match this query, they are delivered automatically to the person who set up the content router (filter). This mode helps people keep up to date with the most recent documents matching their (stable) search interests.

From a linguistic point of view, an interesting and very general conclusion can be drawn from standard IR approaches. A content-based search problem for finding relevant documents is treated as a *lexical* problem (sometimes, when noun phrases are incorporated, it is slightly extended as a phrasal problem). Relevancy, usually, is mirrored through co-occurrence patterns of nouns (as denotations of topics, issues, etc.)—either in sets of documents (Boolean retrieval) or in a vector space. Since all other kinds of text-specific information is lost (even the natural order of words in a text), this approach has been dubbed as a 'bag of words'. Although numerous attempts have been made to promote further linguistic methodologies beyond stemming (e.g. syntactic processing of phrases), their application yielded discouraging results in many experiments. By and large, linguistic methods did not outperform the much simpler statistical ones and failed to improve the performance of IR systems substantially.

While the methods discussed so far cover a traditional, fairly established view on the IR task, viz. the provision of relevant *documents*, current research tries to complement this document-centered view on IR by focusing on relevant *facts* and *data*. A corresponding *information extraction* task (Gaizauskas and Wilks 1998; and Jackson and Moulinier 2002) can then be phrased as follows: given a set of documents, extract all of the information considered relevant from these textual sources by filling a priori supplied sets of domain-specific templates (reminiscent of the knowledge representation construct of frames). Typically, these templates address names of people, locations or organizations, as well as simple relations and complex events involving these entities. In order to achieve this goal, a much more sophisticated analytic machinery than the one used for classical IR tasks has to be provided. Usually, it is based on regular expressions for named entity recognition, (cascades of) finite-state automata for partial syntactic analysis (parsing), template-filling rules which move linguistically analyzed text portions (e.g. noun or verb groups) into the canonical template structures, while template merging rules account for the final assembly of consistent and complete data about the same entity or event as captured from the entire document. Also in hybrid approaches, statistical methods play their part, although not as dominantly as in the pure IR framework.

Text mining (Hearst 1999) is slightly different from the information extraction task. Although the target is the extraction of factoid knowledge from textual sources, for text mining there are no a priori templates available, which specify the type of relevant information a system should look for. Rather, text miners try to capture 'new' and interesting (relevant) information for which no a priori conceptualization exists yet. By way of exploring textual data and interpreting it, a text miner supplies its user with unexpected and, thus, particularly valuable information. As with classical document-focused IR, the notion of relevance and the interest value is crucial for even more advanced fact-finding IR activities.

Another challenging task is concerned with factual *question answering* directly from large collections of on-line textual resources (Harabogiu and Moldovan 2003). Unlike information extraction whose scope is limited to the set of a priori available templates, text-based question answering is open-domain, i.e. not constrained to preselected topics. Successful question answering systems use pretty advanced natural language processing techniques, e.g. to capture the semantics of (*wh*) questions and to perform lexico-semantic unifications of questions and candidate answers (usually, text snippets) through the collection and assessment of evidence from multiple documents.

Recently, the field has further expanded into the area of *text summarization*, although this term is a bit of an overstatement considering the methods being used (Mani 2001). Basically, a summary is an abreviated version of a source text, which contains the most relevant information from its source, with compression ratios ranging from 5 to 25%. Single-document summarization can even be considered a special case of *multi*-document summarization where the contents of a whole set of thematically homogeneous documents has to be transformed into one comprehensive summary (redundancy and overlap determination, resolution of inconsistencies, and temporal sequence errors then become crucial issues).

Since in-depth abstraction from the source has proved to be infeasible so far, the problem has been rephrased as a sentence extraction problem. An *extract* is then formed from sentences, which are likely to contain the gist of or crucial statements of the source document. Few types of surface indicators seem to be sufficient. Given the prevailing text genres in IR (newspaper articles and technical reports), it is no surprise that cue phrases indicating the salience of a sentence (such as 'in summary', 'our experiments have shown', etc.) or the position in which a sentence occurs (e.g. lead sentences in a journal article, the first and last paragraph of a document) are really good indicators for picking up relevant sentences for an extract, while statistically relevant terms (e.g. based on some of the statistical measures mentioned for the classical

IR approach) are less reliable but still useful. Most extraction systems vary in the degree they tune the particulars of these parameters, manually or even automatically (when considered as a machine learning task).

The field of IR has a strong experimental backing. Given soft notions such as relevance, salience, importance, which are crucial for almost all IR tasks, the accomplishment of how relevant the delivered information items really are needs quantitative assessment. Among the various quality metrics discussed in the field, two particularly relevant ones have found considerable consensus, viz. precision and recall (cf. Table 1).

The f-measure (cf. Table 2) is a convenient way to combine the effects of recall and precision, allowing the designer to stress the impact of either one of these (via the weights α and β) depending on the experimental setting.

While all the above-mentioned tasks deal with *textual* data only, a real challenge to IR will be the increasing amount of *multimedia* data. This includes audio (speech, sounds, music), graphics (tables, formulae, graphs, images, animation), and video data (movies),

TABLE 1   Precision and Recall Metrics

Precision	:=	$\dfrac{\#relevant\ items\ retrieved}{\#items\ retrieved}$
Recall	:=	$\dfrac{\#relevant\ items\ retrieved}{\#relevant\ items\ in\ the\ collection}$

TABLE 2   F-Measure

F-measure := $\dfrac{(\alpha \times Recall) \times (\beta \times Precision)}{(\alpha \times Recall) + (\beta \times Precision)}$

for which only modest progress has been made so far. With the WWW becoming the prime information exchange platform for information seekers, additional linguistic challenges come up associated with the multilinguality of the Web. In this area, a distinction is made between *cross-language retrieval* (a native-language query is processed on many other languages, too, although the results are original documents in various foreign languages) and *multilingual retrieval* (in this case, hits are, usually roughly, translated to empower the searcher to assess the relevance of a document and, possibly, arrange for a detailed high-quality translation, if the document is really relevant).

References

Baeza-Yates, Ricardo, and Berthier Ribeiro-Neto (eds.) 1999. Modern information retrieval. Reading, MA: Addison-Wesley & Longman.

Gaizauskas, Robert, and Yorick Wilks. 1998. Information extraction: beyond document retrieval. Journal of Documentation 54(1). 70–105.

Harabagiu, Sanda, and Dan Moldovan. 2003. Question answering. The Oxford handbook of computational linguistics, ed. by Ruslan Mitkov, 560–82. Oxford, UK: Oxford University Press.

Hearst, Marti A. 1999. Untangling text data mining. Proceedings of the 37th annual meeting of the Association for Computational Linguistics, 3–10. College Park, MD, USA, 20–26 June 1999. San Francisco, CA: Morgan Kaufmann.

Jackson, Peter, and Isabelle Moulinier. 2002. Natural language processing for online applications. Text retrieval, extraction and categorization. Natural Language Processing, Vol. 5. Amsterdam, Philadelphia, PA: John Benjamins.

Mani, Inderjeet. 2001. Automatic summarization, Natural language processing, Vol. 3. Amsterdam, Philadelphia, PA: John Benjamins.

van Rijsbergen, C.J. 1979. Information retrieval, 2nd edition. London, UK; Boston, MA: Butterworths.

Salton, Gerard, and Michael J. McGill. 1983. Introduction to modern information retrieval. New York, NY: McGraw Hill.

<div align="right">Udo Hahn</div>

Interpreting

In interpreting, a spoken message or text in the original ('source') language is converted into a spoken message or text in another ('target' or 'receptor') language. Interpreting and written translation have certain similarities, but the on-line, more directly interactive nature of most spoken text gives spoken interpreting specific features.

Three key variables in interpreting are time lag, physical/professional setting, and language direction. Looking first at time lag, the crucial difference is between consecutive and simultaneous interpreting.

In consecutive mode, the interpreter hears the source text or a section of it (e.g. one speaker's turn, as with Message 1 in Figure 1); the source speaker then stops
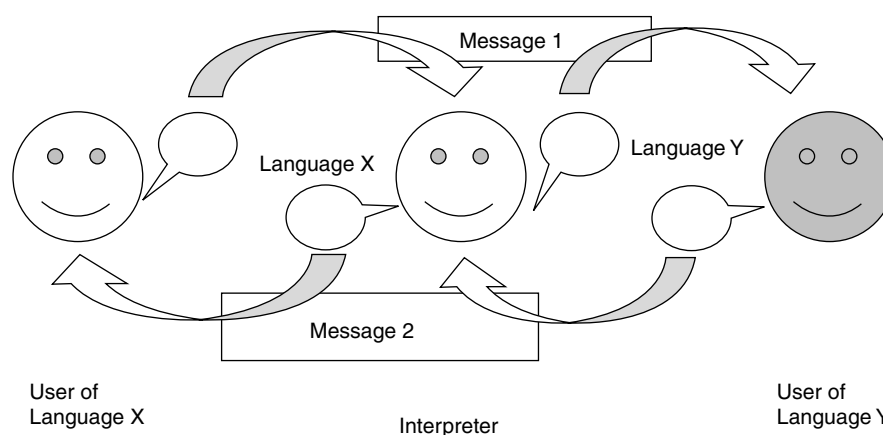
Figure 1

speaking, and the interpreter delivers a translated version to the target listener. While listening, interpreters can use strategies such as note-taking and mnemonics to remember source text data, and they have time to work out the text structure (e.g. whether the first sentence contains a preamble or the main point). Good interpreters will pass on all information contained in the source message, but they are unlikely to produce a 'literal' translation. Just as 'normal' listeners, after a certain delay, remember the meaning of a message rather than exact wording or grammar, consecutive interpreters use memory of meaning to construct the target message; thus, 'places like Thailand, Malaysia, Indonesia, and Vietnam' might be translated as 'Southeast Asian countries' (see Figure 1).

In simultaneous interpreting, interpreters produce a target text while the source text is coming in: once they have received a translatable source-text segment, they pass it on in the target language while listening to the next source-text segment. The translatable segment is typically a meaning-based unit a couple of phrases long: e.g. '// the government maintains that // BSE, or 'mad cow disease', // is of no danger to the public. //' Simultaneous interpreting thus requires two closely related subskills: divided attention, i.e. the ability to pay attention to both the incoming and the outgoing message, and multitasking, i.e. the ability to perform several tasks simultaneously. The key tasks to be done simultaneously (see Figure 2) are as follows:

- Analyzing and comprehending the incoming source-language segment, fitting its information into an emerging picture of sentence and overall source-text meaning, discourse structure, speaker attitude, etc., and holding the meaning in short-term memory.
- Planning and producing the target-language segment according to the previous source-text segment and to the emerging sentence and whole-text structure.

- Monitoring (closely scrutinizing) the emerging target text and, if necessary, repairing it (correcting, e.g. inserting 'I'm sorry, that should have been X').
- Strategies can be used to make the task a little easier. For example,
- Several segments may need to come in before one has enough information to start translating (e.g. because the source and target languages have very different sentence structures, such as English and Japanese): then, when the interpreter can finally deliver a target version, he or she does so as quickly as possible to clear working memory.
- Exploit the source speaker's pauses to deliver as much target text as possible without having to process new input.

Nevertheless, simultaneous interpreting remains a specialized and taxing skill. Becoming a simultaneous interpreter requires extensive training and practice, and even experienced interpreters cannot interpret for so long at a time (typically, 20 minutes between breaks: hence, conference interpreters often operate in two-person teams). Preparation, however, is crucial in easing the difficulty of both consecutive and simultaneous interpreting: thus, before an assignment, interpreters usually ask for documentation, drafts of speeches, etc., to be given to them in advance.

A second key variable in interpreting is the physical setting. The simultaneous interpreter is usually in a soundproofed booth, where he or she hears the source speaker through headphones and produces the target message into a microphone. Conference delegates, courtroom advocates, etc., can listen either to the interpreter through their own headphones (at conferences, switch settings may give a choice of target languages) or to the source speaker directly. With just one target-language listener, simultaneous interpreting may also
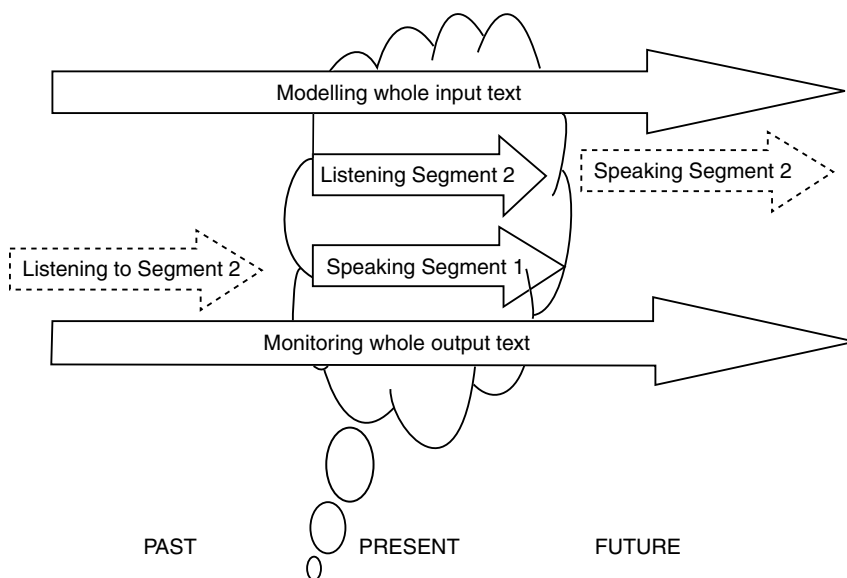
Figure 2
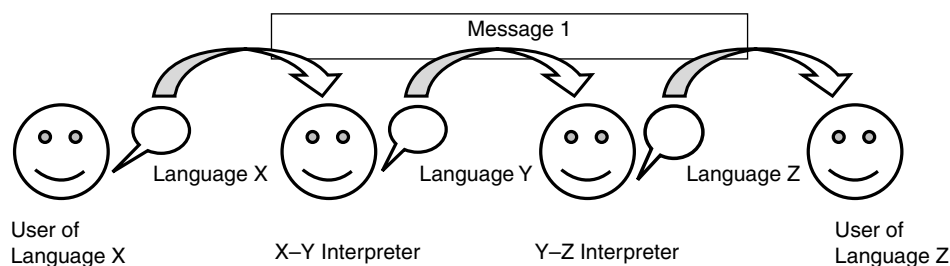
Figure 3

occur through *chuchotage*, where the interpreter sits close to the listener, hears the source message directly and whispers the target version into his or her ear.

In 'liaison interpreting', a consecutive interpreter enables communication between two parties—as in business or in 'community' interpreting (working with members of linguistic minorities at hospitals, with immigration authorities, etc.). The interpreter is physically present, often positioned between the two parties, and thus tends to be regarded as a fuller participant in the interaction. Usually, there is only one interpreter (see Figure 1), although sometimes—e.g. in international political talks—there may be two (one for each language direction).

In most settings—courtroom interpreting, for example—accurate rendering of source text and strict neutrality between parties are key professional ethics. In some settings (e.g. when mediating between members of different community groups, or in business negotiations), however, interpreters may see their job as facilitating communication in a wider sense. This may involve, for instance, explaining the cultural assumptions behind one party's behavior—e.g. telling a bereaved American that an Indonesian's smile indicates sympathy rather than amusement.

As for language direction, interpreters usually prefer to translate from their 'B language' (a nonnative language they know well) into their 'A language' (the language they know best—usually their native language). Research, however, indicates little difference between A→B and B→A performance with skilled interpreters (in the former, less attention is needed for source-text comprehension, and in the latter, less attention is needed for target-text production). Moreover, in some language pairs, nearly all interpreters are native speakers of one language rather than the other (e.g. most Bosnian–English interpreters are native speakers of Bosnian rather than English) and thus often find themselves having to work in both directions.

Sometimes, interpreters may not be available for a certain language pair (e.g. at a conference with English, Spanish, and Japanese as working languages, there may be no Spanish-Japanese interpreters). Then, 'relay interpreting' may take place, for

example, a Spanish speech is interpreted into English, and the English version is interpreted into Japanese; see Figure 3.

References

Carr, Silvana E. (ed.) 1997. The critical link: interpreters in the community. Amsterdam and Philadelphia: Benjamins.

Gambier, Yves, Daniel Gile, and Christopher Taylor. 1997. Conference interpreting: current trends in research. Amsterdam: Benjamins.

Gile, Daniel. 1995. Basic concepts and models for interpreter and translator training. Amsterdam: Benjamins.

Jones, Roderick. 1998. Conference interpreting explained. Manchester: St. Jerome.

Lambert, Sylvie, and Barbara Moser-Mercer (eds.) 1994. Bridging the gap: empirical research in simultaneous interpretation. Amsterdam: Benjamins.

Pöchhacker, Franz, and Miriam Shlesinger (eds.) 2002. The interpreting studies reader. London: Routledge.

Wadensjö, Cecilia. 1998. Interpreting as interaction. London: Longman.

FRANCIS R. JONES

Irony

The word *irony* comes from the Greek *eironeia* ('pretense, dissimulation'), as does the history of its definition and analysis. Irony is seen as a trope (i.e. a figure of speech) in ancient rhetorics, and this analysis has remained essentially unchallenged until recently. In the traditional definition, irony is seen as saying the opposite of what one means. This definition is demonstrably incorrect, because a speaker may be ironic but not mean the opposite of what he or she says, as in 'It seems to be a little windy' (uttered in the middle of a violent storm), in which the speaker is saying less than what is meant. A recent and fruitful restatement of the irony-as-trope theory has been presented by H. Paul Grice, who sees irony as a linguistic structure with only implied meaning, which puts a significant burden on the listener, who has to decode the implication. Broadening the definition to, for example, 'saying something while meaning something else', runs the risk of obliterating the specific difference between irony and other forms of figurative or indirect speech. Despite the problems with the idea of 'oppositeness', approaches to irony as negation have been presented. Speech-act approaches to irony see it as an insincere speech act. Other approaches to irony include the 'tinge' theory, which sees irony as blending the two meanings (the stated and the implied ones) with the effect of attenuating the ironic one.

A very influential approach to irony is the mention theory (Sperber and Wilson 1981), which claims that an utterance is ironic if it is recognized as the echoic mention of another utterance by a more or less clearly identified other speaker. Furthermore, the ironic statement is critical of the echoed utterance. Similar theories based on the ideas of 'pretense' and 'reminder' have been presented as well. Criticism of the mention

theory notes that not all irony seems to be interpretable as the echo of someone else's words and that an admittedly rarer, nonnegative, praising irony exists. An example of the latter is the utterance 'Sorry to bother you' before the announcement of excellent news (e.g. winning the lottery).

All the theories of irony mentioned so far (except mention theory, which is presented as a direct access theory; see below) share the idea that the processing of irony is a two-step process in which one sense (usually assumed to be the literal meaning) of the utterance is accessed and then a second sense of the utterance is discovered (usually under contextual pressure). Thus, in a Gricean account of irony as implicature, the hearer of an utterance such as 'That was smart' (uttered as a description of clumsy behavior, such as spilling one's wine on someone's clothing) will first process the utterance as meaning literally roughly 'This behavior was consonant with how smart people behave.' This interpretation will then be discarded in favor of the implicature that the speaker means that the behavior was *not* consonant with how smart people behave. This account has been challenged recently by 'direct access' theories.

The direct access theories claim that the hearer does not process the literal meaning of an ironic utterance first and access only later the figurative (ironic) meaning. They claim instead that the literal meaning is accessed either not at all or only later. Direct access interpretations of irony are directly at odds with the traditional interpretation of irony as an implicature. Some results in psycholinguistics have been seen as supporting this view (Gibbs 1994). The mention theory of irony is a direct access theory. Other researchers have presented contrasting views that support the two-step

approach, although not always the claim that the literal meaning is processed first: claims that interpretations are accessed in order of saliency or in parallel have been put forth.

Psycholinguistic studies of irony have focused on children's acquisition of irony, progressively lowering the age at which they understand it to under 10 years. These studies focus on the neurobiology of the processing of irony, emphasizing the role of the right hemisphere alongside the left one, in which most language processing takes place, and the order of activation of the various meanings in the ironic text.

Considerable attention has been paid to the optional markers of irony, i.e. indications of the speaker's ironic intent through, primarily, intonation and body language. Although several phonological (sound) features have been considered 'markers' of irony, it seems that none is exclusively a marker of irony.

Recently, the social and situational context of irony, as well as its pragmatic ends, have begun being investigated in sociolinguistics and discourse/conversation analysis, and also in psycholinguistics. Work on the social functions of irony has found a broad range of functions, including in- and out-group definition, evaluation, aggression, politeness, verbal play, and many others. It is possible that the list may be open ended. Work on the reception of irony by its audience is also beginning to appear.

It should be noted that the term *irony* is also commonly used to describe 'situational' irony (i.e. irony of facts and things entirely dissociated from their linguistic expression), such as a fire station burning to the ground, Socratic irony, romantic irony, and even a type of religious experience (Kierkegaard). Although there may be connections between situational and verbal irony, it does not seem that literary and religious uses can be fruitfully explained in terms of linguistic irony. Other definitional problems include the purported distinction between irony and sarcasm. Although some have argued that the two can be distinguished (for example, irony can be involuntary, whereas sarcasm cannot), others maintain that no clear boundary exists. Another unresolved issue is the connection between irony and humor. It is clear that the two overlap significantly, but also that they differ: not all irony is humorous, and not all humor is ironic.

References

Attardo, Salvatore. 2000. Irony as relevant inappropriateness. Journal of Pragmatics 32(6). 793–826.

Booth, Wayne. 1974. A rhetoric of irony. Chicago: University of Chicago Press.

Colston, Herbert L. 1997. Salting a wound or sugaring a pill: the pragmatic function of ironic criticism. Discourse Processes 23. 25–45.

Dews, Shelley, and Ellen Winner. 1999. Obligatory processing of literal and non-literal meanings in verbal irony. Journal of Pragmatics 31(12). 1579–99.

Gibbs, Raymond W. 1994. The poetics of mind: figurative thought, language, and understanding. Cambridge, New York: Cambridge University Press.

Giora, Rachel. 2003. On our mind. Salience, context, and figurative language. Oxford: Oxford University Press.

Haiman, John. 1998. Talk is cheap: sarcasm, alienation, and the evolution of language. Oxford, New York: Oxford University Press.

Katz, Albert N. (ed.) 2000. The uses and processing of irony and sarcasm. Special issue of Metaphor and Symbol 15(1/2). 1–107.

Kreutz, Roger, and Richard M. Roberts. 1995. Two cues for verbal irony: hyperbole and the ironic tone of voice. Metaphor and symbolic activity 10. 21–31.

Sperber, Dan, and Deirdre Wilson. 1981. Irony and the use—mention distinction. Radical pragmatics, ed. by Peter Cole. 295–318. New York, London: Academic Press.

Toplak, Maggie, and Albert N. Katz. 2000. On the uses of sarcastic irony. Journal of Pragmatics 32(10). 1467–88.

Winner, Ellen. 1988. The point of words: children's understanding of metaphor and irony. Cambridge, MA: Harvard University Press.

SALVATORE ATTARDO

See also **Grice, H. Paul**

Israel

One does not ordinarily speak of 'miracles' in linguistics, and with good reason. However, if any event in linguistics and language history deserves the name 'miracle', it is the rebirth of the Hebrew language in modern Israel. Languages, once dead, do not normally rise again from the dead. This the Hebrew language did.

True, Hebrew was never completely dead; it is more accurate to speak of the 'revival' of the Hebrew language rather than its rebirth. Rabbis and learned men never ceased to use Hebrew for high purposes long after it had declined as the spoken language of the Jewish people; it was always the language of prayer

for observant Jews; it served to some extent as a lingua franca among Jews from different countries. But for over a millennium and a half until toward the end of the nineteenth century, Hebrew was virtually nobody's first language. It was a language acquired in adolescence and adulthood by at most a minority of Jewish males, by Jewish females almost never because of religious strictures. Today, it is a language like any other—that is to say, a language spoken by ordinary people going about their everyday activities and acquired in the usual way: by learning the language in infancy.

Hebrew is a member of the Semitic family of languages, to which Arabic and Amharic and extinct languages such as Aramaic and Phoenician also belong. It is written from right to left in what is essentially a consonantal script. Semitic languages are organized around the principle of the root: roots have a basic meaning expressed by means of, usually, three consonants. The Semitic root *SLM* elucidates the principle. Its basic, abstract meaning is 'peace', and from it are derived such words as Arabic *SaLaaM*—the greeting (in essence 'peace be with you', cf. Hebrew *ShaLoM*. Consider also *iSLaM* and (m)*uSLiM* and the proper names from Jewish history such as (Hebrew) *SoLoM*(*on*) and *SaLoMe* or (Arabic) *SuLeiM*(*an*). The reader of Arabic or Hebrew recognizes that the basic 'meaning' of the word or name is carried by the root structure *SLM*. The common semantic intersection of these and other words formed with the three consonants *SLM* is 'peace'.

Normally, Semitic words are written without vowels. This is possible because the context suggests the range of readings of a given word in print and because there are rules of grammar that help the reader know what vowel is to be supplied. In Hebrew, for example, the word *spr* consisting of three letters can be read /safar/ 'he counted', /siper/ 'he told', /supar/ 'it was told', /sapar/ 'barber', /sefer/ 'book', and /sfar/ 'frontier'. (In Hebrew, the same letter is used for /p/ as for /f/.) 'Full spelling', in which diacritic marks are used to indicate the vowel, can be used in Hebrew but normally is not. The reader simply has to know from context and native linguistic knowledge of the language which reading is the correct one. In Hebrew, vowel marking is used principally in pedagogical materials, children's books, Biblical texts, poetry, and dictionaries.

Israel, earlier called Palestine, has always been an area of intense historical and linguistic collisions with aftershocks extending far beyond the tiny piece of soil that the country actually encompasses. It was the birthplace of two major world religions, Judaism and Christianity, always a country pulled this way and that and culturally enriched successively by Jews, Greeks,

Romans, and Arabs, only to name the principal players who came on and off the Palestinian stage during its long history. The languages of all its occupiers contributed to a multilinguistic profusion and vibrancy that is preserved in the linguistic pluralism of present-day Israel.

Hebrew was the common language of the Jewish people—the everyday spoken language of the people of the kingdom of Judah—from around 1000 BCE and thus is at least 3,000 years old. It was the language in which most of the Jewish Bible (called in Christian usage the Old Testament) was written (the 'New Testament' of Christianity was written in Greek, not Hebrew). However, another Semitic language, Aramaic, was used for legal and official purposes by priests and officers of the court. In 800 BCE, most of the Jews of Palestine were forced into exile—the Babylonian Exile. Seventy years later, when Jews were permitted to return to their ancestral land and most did, the linguistic situation in Palestine had begun to shift dramatically. Aramaic had descended from the court to the marketplace, and it developed rapidly among both Jews and non-Jews as the major vernacular language in the Middle East, a position it maintained until well into the Common Era when Arabic took its place. Jews for the most part became bilingual in Hebrew and Aramaic, and from this time forward the days of Hebrew as the spoken language of everyday life were numbered. Jesus of Nazareth spoke Palestinian Aramaic as his native tongue.

When the Holy Land came under Greek dominion, the Jewish upper classes increasingly became 'Hellenized', meaning that they adopted Greek cultural practices and the Greek language. As Palestine passed from Greek into Roman hands, Latin also became a language that upwardly mobile Jews had to learn. Thus, by the beginning of the Common Era, bilingualism (Aramaic/Greek or Aramaic/Hebrew) was common, trilingualism (Aramaic/Greek/Hebrew or Aramaic/Greek/Latin) not uncommon, and the Jerusalem Talmud—the Talmud is the body of authoritative Jewish tradition and commentary comprising the Mishnah and Gemara—states that 'four languages are of value: Greek for song, Latin for war, Aramaic for dirges, and Hebrew for speaking'. Monolingual Hebrew speakers carried on only in remote villages in southern Judea. By the second century CE, Hebrew had become so precarious as a spoken language that we find rabbinical injunctions that fathers should speak Hebrew with their sons, a certain sign that most fathers were not speaking Hebrew with their sons.

Although Hebrew receded as spoken vernacular, it retained its significance as the language of the Bible—holy, sacred, divine—and both Aramaic and Hebrew are joined in the Hebrew designation *leshon hakodesh*,

meaning 'the language of sanctity'. To this day, Aramaic is the language in traditional Judaism of legal instruments such as a divorce decree (*get*) and the marriage contract (*ketuba*). Bilingualism in Hebrew and Aramaic is implicit in the transmission of traditional Judaism.

What is known as the Jewish Diaspora began with the destruction by the Romans of the Second Temple in 70 CE. As Jews spread out from their ancestral homeland into Europe and the lands bordering the Mediterranean, like all emigrants to foreign lands, they began to speak the language of the country in which they had settled during the early and medieval periods of our millennium. Jews in Spain spoke medieval Spanish—perhaps an identifiably Jewish variant of Spanish in some cases, but Spanish nevertheless. Jews in France spoke Old French or, again, possibly a Jewish dialect of Old French. Jews living along the Rhine and Danube rivers in Germany spoke Middle High German—the German of 1050 to 1350 CE—of one kind or another, probably always with a Jewish flavor and accent.

When the Jews were expelled from Spain in 1492 they settled for the most part around the Mediterranean Sea—in northern Africa, Italy, Greece, Turkey—but the language they carried with them was Spanish. In time, their Spanish changed and developed into the language now usually called Ladino (the usage is not settled: one also finds 'Judeo-Spanish', 'Judezmo', 'Dzhudezhmo', and 'Spaniolish'): it is Spanish with admixtures from Hebrew and the languages of the lands to which the Spanish emigrants had fled. Many of these Jews gave up their ancestral Spanish altogether and acquired the language of the country to which the winds had blown them. This branch of the Jewish people is called the *Sephardim* from the medieval Hebrew word for 'Spain'.

A different set of language choices obtained among the Jews, who had settled during the Diaspora from the Holy Land in western Europe north of the Pyrenees, primarily in Germany. These Jews are called the *Ashkenazim*, after the medieval Hebrew word for 'Germany'. Until the eleventh century CE, Jews had lived in France, Germany, and England in relative peace and security. They were technically, if often only notionally, under the protection of the Holy Roman Emperor. This changed terribly for the worse with the onset of the Crusades in 1096. What began as a war against the 'Saracens'—'Arabs', 'Turks', later 'Moors'—to reclaim the holy sites of Christianity rapidly became a war against 'unbelievers' of every stripe, and hence Jews were expropriated and massacred and expelled until finally there was nothing left for them but wholesale emigration away from the troubles. In practical terms, this meant fleeing to Poland and other countries of central and eastern Europe, where there was at the time little anti-Jewish sentiment and where Jews were welcomed for economic reasons. There was no sanctuary in the west. The language the fleeing Jews carried with them was medieval German, which on the soil of eastern Europe and in isolation from the German language in Germany, developed into the language for which the preferred designation is Yiddish (earlier 'Judeo-German').

Yiddish is written in the Hebrew alphabet, like Ladino and virtually all Jewish languages, but is some 80% German in vocabulary, 15% Hebrew, and 5% Slavic and other languages. The marginal vocabulary of Yiddish depends on where the Yiddish language is spoken—American Yiddish has borrowings from American English, French Yiddish from French. However, the basic language structure of Yiddish is Germanic through and through.

Yiddish is one of the many languages that have always suffered from inferiority complexes. Because, although it shared ancestry with German, it had developed in ways different from the standard German language, Yiddish was condemned by some as 'bad German'. Others argued, especially with the spread of Zionism in the late nineteenth and early twentieth centuries, that Jews should return to their ancient ancestral language, Hebrew, that only this ancient ancestral language could unite Jews from all parts of the world in a common quest for security and a place of their own. Still others argued for assimilation: Jews living in Poland should speak Polish, Jews living in America English, Jews living in Russia Russian, and so on.

Every place that Jews have lived they have developed 'Jewish' variants of the local language. Thus, we have Yevanic (Judeo-Greek), Dzhidi (Judeo-Persian), Judeo-Georgian, Judeo-Tat, Judeo-Tunisian Arabic, Judeo-Iraqi Arabic, and many others. In the medieval era, there was Laaz (Judeo-French) and Knaanic (Judeo-Slavic). These languages, to the extent that they still are spoken, are generally on the list of endangered languages. Ladino and Yiddish stand out among 'Jewish languages' because they are the two best known with the most extensive literatures and the largest numbers of speakers—at least until the World War II destroyed most of European Jewry. The number of speakers varies widely. In Israel today, there are estimated to be some 200,000 speakers of Yiddish, 100,000 speakers of Ladino, 250,000 speakers of Judeo-Moroccan Arabic, 60,000 speakers of Dzhidi, and 35 speakers of Yevanic. These Jewish 'legacy' languages are in addition to languages still spoken by some Israelis such as Bulgarian, French, German, Hungarian, and English. Very few people would be monolingual in one of the Jewish heritage languages,

almost all would be fluent to some degreee in Hebrew. Fluency in Hebrew has had the highest priority since the creation of modern Israel in 1948.

Hebrew remained dormant through the centuries, its use confined largely to the written form of the language, especially among the rabbinate and in the synagogue. However, it never ceased to be a written language. People whose cradle language had been anything in the world but Hebrew composed substantial works in the language, often elegantly and with originality. Every Jewish boy had to learn Hebrew in the Hebrew alphabet—well or badly, as is true today—in order to become *bar mitzvah*, signifying that he had reached his thirteenth birthday and had, therefore, attained maturity and had reached the age of religious duty and responsibility. Females did not have a coming-of-age ceremony in traditional Judaism and thus were largely illiterate in Hebrew. 'Literacy' is, of course, not nearly the same as 'fluency': to be literate simply meant that religiously observant Jewish males could read Hebrew. Jewish languages like Yiddish and Ladino, indeed virtually all Jewish variants of languages, were and are written in Hebrew characters. This goes back to the widespread literacy of Jews in Hebrew—of Jewish males at least—during the 'Dark Ages'.

The revival of Hebrew as a secular written language began during what is known as the Jewish Enlightenment (*Haskalah*) in the eighteenth century. Talented writers began to write in the language, play with it, revivify it, although they often wrote in Yiddish as well. The revival of Hebrew as a spoken language began in small steps in the late nineteenth century. One of the main reasons, ultimately, was anti-Semitism. Ill treatment of Jews—pogroms, discrimination—was on the increase in eastern Europe and Russia, and even western Europe, where since Napoleonic times the legal situation of Jews had improved considerably, was not immune. The Dreyfus Case in France, in which a Jewish officer, Alfred Dreyfus, was falsely accused of espionage and condemned to a harsh imprisonment on Devil's Island simply because he was Jewish, laid bare how thin the veneer of civility was when it came to being Jewish even in supposedly enlightened western Europe.

The darkening clouds over Jewish life in Europe led to the *Hibbat Zion* movement ('Love of Zion') and to the Zionist movement itself, founded by the Austrian Jewish journalist Theodor Herzl, which called for the creation of a Jewish homeland. Jews began to leave eastern Europe and settle in largely agricultural settlements in Palestine that they had purchased or leased from Arab or Turkish landowners. (Palestine was then part of the Ottoman Turkish Empire.)

The first wave of immigrants arrived between 1882 and 1903 (called the *First Aliyah,* or first 'homecom-

ing'). *Aliyah* brought a renewed interest in spoken Hebrew, and in the newly regained ancestral land, schoolteachers and families here and there began to use spoken Hebrew. Eliezer Ben-Yehuda (1858–1922) was one of the leaders in the movement, but it is a myth that he single-handedly brought about the revival of spoken Hebrew by refusing to speak anything else with his family. The movement was slow in the beginning. Competition from traditional Jewish Diaspora languages such as Yiddish and Ladino was stiff, not to mention French and German in the schools, and it has been estimated that by 1904 there were only about 20 Hebrew-speaking families in Palestine.

The movement accelerated with the *Second Aliyah* (1904–1906). Waves of *po'alim* 'workers' arrived, and they soon had established Hebrew-medium high schools (*gymnasia*) and teachers' seminaries to cater to the growing commitment in the Jewish population of Palestine to the resurrection of spoken Hebrew. The movement fed upon itself. To speak Hebrew was to assert that one belonged to the worldwide community of Jews and was a proud member of that small but growing band of Zionist pioneers who had returned to the ancestral home. The larger the number of Hebrew-speaking students coming out of the school system, the greater the demand for entertainment and reading material in Hebrew—plays, novels, newspapers. By 1916–1918, some 40% (34,000 of 85,000) Jews in Palestine recorded Hebrew as their first language.

The problems of bringing an ancient language up to the complexity of modern life were enormous. Words had to be invented for chicken pox, coffee, opera, electricity, idealists, billiard cure, mailman, and so on. Phrases not found in the Bible had to be devised: all right, I'll do it; make it brief; at your service. What was the correct pronunciation of Hebrew? Sephardic and Ashkenazic traditions differed, for example David vs. Dovid, Shalóm vs. Shólem 'hello'. As a rule, the Sephardic pronunciation prevailed. By the 1920s, a majority of the Jews living in Palestine spoke Hebrew, although this represented a tiny minority of the world's Jewish population, and the only ones who spoke it fluently were the *sabras*, the Israelis who had grown up with it.

After World War II had savagely reduced the world's Jewish population from some 11,000,000 to around 5,000,000, Jews emigrated in large and growing numbers to Palestine, and in 1948 the new country of Israel gained its independence from the British. The question was: what should the language of the new country be? Other things being equal, Yiddish would have been the leading candidate since most of the eastern European Jews who had survived the Holocaust and managed to get to Palestine were speakers of Yiddish. Yiddish, however, suffered from

various disabilities, one of which was that it was stigmatized as the 'language of the ghetto'. It was thought of as a victim's language. No language besides Hebrew could possibly do for the new Israel— not English, not Yiddish, not Ladino. Hebrew linked the Jewish past and the Israeli future as no other language could. Hebrew was a sublime symbol of hope, of aspiration—not only of Jewishness but of a muscular strain of Jewishness that would never allow another Holocaust to happen to its people: *Never Again*! became the rallying cry of modern Jewish pride and militancy. The Hebrew language is its symbol, its icon.

As unique an event in the annals of linguistics as the revival of the Hebrew language was, it is only part of a much larger story of linguistic legacy. If the revival of the Hebrew language in Israel was a miracle, it was part of a greater miracle: the preservation of Jewishness through centuries of discrimination, massacre, and oppression. All of the languages of Israel— Hebrew, Aramaic, Yiddish, Ladino, and all the other 'Judeo-X' variants—are part of the Jewish legacy, and all have contributed to the preservation of Jewish ethnicity from the days when Moses led his people out of Egypt into freedom in The Holy Land.

References

Ben-Rafael, Eliezer. 1994. Language, identity and social division: the case of Israel. Oxford: Clarendon Press.

Bunis, David M. 1975. Problems in Judezmo linguistics. New York: The American Sephardi Federation.

Chomsky, Willliam. 1957. Hebrew, the eternal language. Philadelphia: The Jewish Publication Society of America.

Fellman, Jack. 1973. The revival of a classical tongue: Eliezer ben Yehuda and the modern Hebrew language. The Hague: Mouton.

Fishman, Joshua A. (ed.) 1974. The sociology of language in Israel. International Journal of the Sociology of Language 1. The Hauge: Mouton.

Fishman, Joshua A. (ed.) 1981. Never say die! A thousand years of Yiddish in Jewish life and letters. The Hague: Mouton.

Fishman, Joshua A. (ed.) 1985. Readings in the sociology of Jewish languages. Leiden: E.J. Brill.

Hetzron, Robert. 1987. Hebrew. The World's major languages, ed. by Bernard Comrie. New York: Oxford University Press.

Paper, Herbert H. (ed.) 1978. Jewish languages: themes and variations. Cambridge, MA: Association for Jewish Studies.

Rabin, Haim. 1973. A short history of the Hebrew language. Jerusalem: Jewish Agency.

Spolsky, Bernard, and Elana Shohamy. 1999. The languages of Israel: policy, ideology and practice. Clevedon, Buffalo, Toronto, Sydney: Multilingual Matters Ltd.

Weinreich, Max. 1980. History of the Yiddish language. Chicago: University of Chicago Press.

ROBERT D. KING

Italian

Italian (*Italiano*) is a Romance language spoken by more than 65 million people in Italy (including Sardinia and Sicily), Vatican, Malta and Corsica, Southern Switzerland (where it is one of the four official languages), San Marino, Istria, on the Northeastern shore of the Adriatic Sea, in the former Italian colonies of Libya, Somalia, Eritrea and Ethiopia, in Tunisia, Egypt, and Greece. It is also quite spread in the whole Mediterranean area, and is spoken by the large immigrant communities in the Americas, especially in the United States and Argentina, in Canada, North Africa, and Australia.

The earliest known written materials in Italian date from the tenth century; more precisely, a document in the Archives of Montecassino, dated year 960, contains a whole sentence, repeated four times, that is practically Italian vernacular. The first literary work of length is the *Ritmo Laurenziano* (Laurentian Rhythm), a *cantilena* in praise of a bishop by a Tuscan, of the late twelfth century.

Standard Italian is essentially based on the Tuscan dialect, which was adopted first as the literary and then as the official language for the whole Italian geographic region. The reason for this choice was that in the fourteenth century, Dante Alighieri, Francesco Petrarca, and Giovanni Boccaccio wrote their masterpieces using this Italian dialect; more precisely, the Old Florentine dialect of Tuscan, and their works became genuine models of grammar. Their impact was so strong that, even if later in the nineteenth century the dialect of Rome started gaining considerable prestige, it could never equal the Florentine's one.

Apart from the Standard Italian, there are a number of distinct dialects spread along and across the Italian-speaking territories. Some of them are very close to the national language, such as those from Tuscany, Northern Latium, and Umbria, while other dialects are more remote, such as those in the Northern and Southern Italy, as well as those in the isles.

This diversity is mainly due to the individuality of various parts of the territory, which determined that the distinctiveness of a certain region be preserved, both in the case of pre-Latin population, coming from different areas and speaking different languages, and in the case

of present-day populations living in different parts of the country. Another factor that contributed to the individualism of various geographical dialects was the barbarian invasion in the High Middle Ages, which brought new elements into the already shaped language varieties, and, of course, the fact that the political unification of Italy occurred only late in the nineteenth century.

The main groups of Italian dialects are: the Northern dialects, the Central-Southern dialects, and the Sardinian group—the last one considered by some linguists as an entirely different language. The so-called 'Spezia-Rimini line', a dividing line that runs East–West across the territory, separates the first two groups.

The Northern dialects constitute two major subgroups: the Venetian (also called Venetic) dialects, containing the Venetian proper and the Trentino dialects, and the Gallo-Italian dialects, containing the Piedmont, Liguria, Lombardy, Emilia, and Romagna dialects, all of which display a close affinity to French in their pronunciation and truncated terminations.

The Central-Southern dialect groups include the subgroup of the Tuscan dialects, consisting of the Florentine, Arezzo, Cortona, Pisa, Livorno, Lucca and Siena dialects, the subgroup of central dialects, including the Modern Roman, Lazio, Umbria, Southern Marche, Corsican, and Northern Sardinia dialects, the subgroup of the Southern dialects, further subdivided into the Neapolitan dialects and the Sicilian ones, and the subgroup of the Southern and Southeastern dialects, containing the Abruzzi, Naples, Campania, Lucania, Calabria, Otranto, and Sicily dialects.

The last group lists the most distinct dialects, i.e. Ladin, Sardinian, and Istrian. Around 60% of the population can speak the national language, but most of them usually speak a local variety. And, given the communication problems between different dialects, most of the population is now at least 'bilingual', although the proper term would be 'bidialectal'.

The Italian alphabet is: a b c d e f g h i l m n o p q r s t u v x z; j, k, w, and y are only used in foreign words, but j can also be found in some proper names in old writings.

In general, all letters are pronounced in Italian, with the exception of *h*. The letter *c* is pronounced /k/ when preceding *a*, *o*, and *u*, but /t ʃ/ when it is followed by *e* and *i*. The same rule applies to the letter *g* denoting its voiced counterpart. *G* is pronounced as /g/ before *a*, *o*, and *u*, but as /dʒ/ before *e* and *i*. The consonant cluster *gg* is also pronounced /dʒ/ before *e* and *i*. *Z* and *zz* are generally pronounced either /t ʃ/ or /dʒ/. *Sc* before *e* and *i* is pronounced /ʃ/.

The sound system of Italian is quite similar to that of Latin or Spanish. Italian has five vowels /ʌ, ɛ, ɪ, ɔ, ʊ/ and four semivowels: /j, e, o, w/, as well as 21 consonants: /p, b, m f,v, t, d, n, s, z, ts, l, r, ʃ, tʃ, dʒ, k, g, h, ɲ, ʎ/.

It is possible in Italian to have double consonants. Usually, there is an opposition between single and double consonants, as in: *eco* 'echo' vs. *ecco* 'here is', *rupe* 'rock' vs. *ruppe* 'he broke', where the double consonants are clearly pronounced.

Intonational stress in Italian generally falls on the next to last or third from last syllables, but it may range freely over the last three syllables of the word. The Italian orthography marks in writing the grave (`) and the acute (´) accents, which are usually used to distinguish otherwise similar-looking one-syllable words, e.g. *se* 'if' vs. *sè* 'oneself', *e* 'and' vs. *è* 'he is', and to indicate stress on the final vowel in polysyllabic words, e.g. *perchè* 'because', *così* 'so'. The circumflex accent (^) was reserved to poetry and it was used to mark either the ellipsis of a syllable in polysyllabic words, or the contraction of two adjacent vowels in a word.

The punctuation of Italian is rather simple. The main differences from English are that the dash (–) is commonly used to mark the exchange of dialogue, instead of inverted commas, or marks and additional comment, instead of brackets. Also, in the case of indirect questions, Italian does not use the question mark. The capitals are less used than in English. Italian does not use capitals for the names of the months, for the days of the week, for the names of the languages, for adjectives referring to a nation, or individuals belonging to a nation, e.g. *il popolo italiano* 'the Italian people', *due italiani* 'two Italians'.

As a direct offspring of the Latin language, Italian displays a Latin-based vocabulary. The greater part of the Italian vocabulary (or lexis) is formed by Latin words that are either directly or indirectly inherited from Latin, or borrowed from a certain Italian dialect. However, in the course of centuries, several foreign languages have influenced the Italian vocabulary. These languages are French, Spanish, English, German, and Arabic. The French influence is most obvious especially in the territory of the Gallo-Italian dialects, but it is not restricted to it. French have influenced the Italian language over a very long period, and the French loan words entered both the everyday language—some basic terms as *mangiare* 'to eat' and *giallo* 'yello' are borrowed from French—and the more technical jargons—*approccio* 'approach', *destriere* 'charger'. The Spanish influence imposed terms like *toreador, bolero* and, nowadays, *goleador*. The English influence was noticeable in the eighteenth and nineteenth centuries, but it was much more prominent in the twentieth century. However, given the influence of the nationalist movement in the politics during the Fascist period, many English loans have been replaced with Italian counterparts, while others changed their original meaning. Germanic exerted its influence as a result of the Teutonic conquests, predominantly in the contact area of Lombardy. The Arabic influence was most important in Sicily and Southern

Italy, and it was a result of the Saracenic invasions. Some of the most common Arabic words are *albicocco* 'apricot', *ragazzo* 'boy', and *zecca* 'mint'.

In turn, Italian has exerted its influence on the languages of Western Europe, especially on English, French, Spanish, German. Some of the English words of Italian origin are: *balcony*, *bandit*, *broccoli*, *casino*, *gondola*, *incognito*, *inferno, lava*, *macaroni*, *malaria*, *opera, spaghetti*, *studio*, *umbrella,* and *volcano*. Most of the words imposed in different languages are in the field of literature, art and, of course, music. In this respect, it must be mentioned that the contemporary musical terminology all over the world includes Italian words like *allegro*, *andante*, *aria*, *concerto*, *crescendo*, *libretto*, *maestro*, *oratorio*, *piano, quartet*, *solo*, *sonata*, *soprano*, *tempo*, and *virtuoso* almost.

The Italian grammar is similar to that of the other modern Romance languages.

The nouns have distinct inflected forms for gender and number. Most Italian nouns end in a vowel, while those ending in a consonant are usually borrowings from a foreign source, e.g. *il bar* 'pub'*, lo sport* 'sport'*, il film* 'movie'. Generally, nouns ending in *-o* are masculine, nouns ending in *-a* are feminine, while those ending in *-e* may be either masculine or feminine. The foreign borrowings ending in a consonant are generally considered to be masculine. The nouns change their vowel endings from singular to plural. The masculine nouns have a plural ending *-i,* while the feminine nouns have a plural ending either in *–i,* when their singular form ends in *–e,* or in *–e,* when their singular form ends in *–a*. A number of masculine nouns have two distinct plural forms, one for the masculine gender and one for feminine gender, each of them with its own, distinct meaning: e.g. *il membro* 'member' – *le membra* 'limbs', *i membri* 'members'; *il braccio* 'arm' – *le braccia* 'arms' (of a body), *i bracci* 'arms' (of an object). Some nouns have an irregular feminine form. In this case, it is quite common for some speakers to use only a masculine form of the nouns instead of the irregular feminine forms: *il ginocchio* 'knee' – *le ginocchia* or *i ginocchi* 'knees'.

In Italian, articles and other modifiers may accompany the noun and usually their forms agree with the form of the noun in the categories of gender and number: e.g. *il riso* 'the laughter'*, la moglie* 'the wife'.

The article forms are chosen according to the sound of the initial letters of the word it precedes. In the case of invariable nouns, the article helps to identify the categories of gender and number of the respective nouns: e.g. *il re – i re* 'the king/s', *la gru – le gru* 'the crane/s'. Also, in the case of nouns of different gender ending in the same vowel *–e*, the article reveals the gender of the noun in question: e.g. *il monte* 'mountain' (m.) vs. *la mente* 'mind' (f.). Unlike in other languages, the plural forms of the indefinite article are rarely used. The arti-

cles can be combined with some simple prepositions, resulting in contracted forms of the definite article. The combination of the definite article with a simple preposition is labeled 'the combined preposition', e.g. *il + a* : *al, gli + di* : *degli, la + con* : *colla, le + in* : *nelle.*

In Italian, there are two groups of adjectives: those ending in the vowel *–o*, e.g. *severo,* and those ending in *–e*, e.g. *grande*. The adjective also agrees in gender and number with the noun it modifies, but it does not have the same ending: e.g. *la ragazza forte – le ragazze forti* 'the strong girl/s'. *Più* 'more', *meno* 'less' and the suffix *-issimo* are used to form the comparative and the superlative degrees: e.g. *più bello* 'more beautiful', *il più bello* 'the most beautiful', *bellissimo* 'very beautiful', *meno bello* 'less beautiful', and *il meno bello* 'the least beautiful'.

The forms of the pronouns differ according to the category of case. The Personal Pronoun for the first-person singular, for example, has different forms according to the case: *io* 'I' in the Nominative (subject) changes to *me* or *mi* '(to) me' in the Dative (indirect object) and Accusative (direct object), according to whether the pronoun is stressed or unstressed. As for their use, subject pronouns are omitted in everyday speech, and they appear when needed for emphasis or contrast. The stressed form of direct object pronouns, e.g. *me* '(to) me', *te* '(to) you', *lui* '(to) him', *lei* '(to) her' is also used for emphasis or contrast; otherwise, the unstressed forms are used, e.g. *mi, ti, lo, la.*

The possessive adjectives and pronouns have different forms to distinguish whether the possessor is the first, second, or third person, e.g. *mio* 'my/mine', *tuo* 'your/yours', *suo* 'his-her/his-hers', whether the possessor is singular or plural, e.g. *mio* 'my/mine', *nostro* 'our/ours', and, moreover, whether the possessed entity is masculine, e.g. *mio,* or feminine, e.g. *mia,* singular, e.g. *mio*, or plural, e.g. *miei*. Therefore, the possessive pronoun form *vostre* 'yours', for example, displays the following information: the possessor is a second-person plural (as 'you plural'), and the possessed entities are plural and feminine (as 'things' feminine).

The verbs are traditionally classified into three classes called 'conjugations', according to their ending in the present infinitive form: the verbs ending in *–are* belong to the first conjugation, e.g. *comprare* 'to buy', those ending in *–ere* belong to the second conjugation, e.g. *credere* 'to believe', and those ending in *–ire* belong to the third conjugation, e.g. *dormire* 'to sleep'. There are seven moods in Italian, namely Indicative, Conditional, Subjunctive, Imperative, Gerund, Participle, and Infinitive, all of them having similar functions as in English. Some of these moods have additional tenses that express different time relations, such as 'past historic', which is used for completed actions that are not related to the present, and 'imperfect', which is used for

continuous action during which some other action occurs. Some of the tenses are compound, i.e. they are made up of an additional verb, called 'auxiliary', and a specific form of the verb in question (the main verb). In general, the transitive verbs (i.e. those that take a direct object) take the auxiliary *avere* 'to have', and the intransitive (i.e. those that do not take a direct object) and reflexive verbs (i.e. the verbs with a reflexive pronoun) take the auxiliary *essere* 'to be'.

The adverbs are invariable parts of speech. Most of them are formed by adding the suffix *-mente* to the feminine singular form of the corresponding adjectives, e.g. *rapidamente* 'quickly'. While the comparative form of the adverbs follows the rules of the adjectives (see above), e.g. *più gentilmente* 'kinder way', the superlative form is always formed with *molto* plus the adverb e.g. *molto gentilmente* 'very kind way'. The adverb precedes the adjective or the adverb and follows the verb it modifies, e.g. *ha finito presto* 'he finished early'.

In the case of numerals, cardinal numerals indicate the quantity; they are invariable in form and (apart for *uno* 'one') they all go with plural nouns, e.g. *cinque studentesse* 'five students'. The Italian system of writing uses points to separate thousands and commas to separate decimals. When expressing the day of the month, except for the first day, the date is indicated by cardinal numerals, e.g. *nove marzo* 'the 9th of March'. Ordinal numerals indicate a sequence; they agree in gender and number with the noun to whom they refer. The cardinal numeral *uno* 'one' is used like the indefinite article *un, una, uno* 'a/an' and agrees with the noun to whom it refers in gender.

Suffixes play an important role in the Italian vocabulary. Italian frequently uses affective suffixes, usually related to size or emotions. In order to indicate smallness or express affection, the following suffixes are used: *-ino/a/i/e,* e.g. *mamma* 'mother' – *mammina, -etto/a/i/e,* e.g. *libretto* 'little book', *-ello/a/i/e,* e.g. *paesello* 'little village', and *-uccio, -uccia, -ucci, -ucce,* e.g. *caruccio* 'quite expensive'. On the other hand, to denote largeness, the suffixes used are *-one/-ona, -oni/-*

one (plural): e.g. *librone* 'big book'. To convey the idea of a bad or ugly quality, the following suffixes are added: *-accio, -accia, -acci, -acce,* e.g. *ragazzaccio* 'rough boy'. To form collective nouns, the following suffixes are used: *-ame, -ume, -aglia,* the latter two also having a derogatory connotation, e.g. *pollame* 'poultry', *forestierume* 'foreign scum', *antiglia* 'old junk'. The suffix *–one* is used to denote specific agents, e.g. *imbroglione* 'cheat', *beone* 'drinker', while the suffix *–io* indicates repetitive tasks or actions, e.g. *lavorio* 'constant working', *mormorio* 'murmuring'. The suffix *–ardo* is used to form pejorative words, e.g. *bastardo* 'bastard', *vecchiardo* 'nasty old man'. Some of the suffixes, like *–accio* may also be used as independent words, e.g. *proprio accio* 'really bad'. Some of the suffixes also combine with each other, e.g. *–etto + -ino*: *librettino* 'little book', *-one + -ino*: *cartoncino* 'small card', *-etto + -accio*: *cagnettaccio* 'nasty little dog'. The suffixes may also be added to verbs, e.g. *lavorare* 'to work' – *lavoricchiare* 'to work a little', *dormire* 'to sleep' – *dormicchiare* 'to snooze', to adverbs, e.g. *beno* 'well' - *benino* 'quite well', *presto* 'early' – *prestino* 'quite early', and to personal names: e.g. *Gigi – Gigino, Peppe – Peppino, Simona – Simonetta.*

Referenes

Aust, Derek, and Mike Zollo. 1997. Azione grammatica: a new Italian grammar. London: Hodder & Stoughton.
Clivio, Gianrenzo P., and Marcel Danesi. 2000. The sounds, forms, and uses of Italian: an introduction to Italian linguistics. Toronto: University of Toronto Press.
Dardano, Maurizio, and Pietro Trifone. 1997. La nuova grammatica della lingua italiana (The new grammar of the Italian language). Bologna: Zanichelli.
Haller, Hermann W. 1999. The other Italy: the literary canon in dialect. Toronto: University of Toronto Press.
Ledgeway, Adam. 2000. Comparative syntax of the dialects of Southern Italy : a minimalist approach. Blackwell Publishers.
Lepschy, Anna Laura, and Giulio Lepschy. 1988. The Italian language today. London: Hutchinson.
Proudfoot, Anna, and Francesco Cardo. 1996. Modern Italian grammar: a practical guide. London: Routledge.

RADU DANILIUC

Italy

Italy, officially known as the Italian Republic (*Repubblica italiana*), is a country situated in southern Europe. It consists of the Italian Peninsula, two larger isles (Sicily and Sardinia), and quite a number of smaller ones. It has an area of 301,338 square kilome-

ters, counts 57,679,955 inhabitants and is divided into 20 regions (*regioni*).

The country's official language is Italian, which is a Romance language, i.e. a language developed from the spoken variant of Latin (other Romance languages are

French, Spanish, Portuguese, Romanian, etc.). However, the country is characterized by a large linguistic diversity. This diversity includes various dialects and minority languages. It is important to note that, in the case of Italy, the term 'dialect' is not intended in the way it usually is: it indicates a local variety that has rather little in common with the standard language and with other Italian 'dialects' (Lepschy-Lepschy). Actually, these varieties should be considered as an outcome of different developments from spoken Latin just like standard Italian, French, Spanish, and other Romance languages.

An explanation for this peculiar situation can be found in history: during centuries, the Italic Peninsula was a patchwork of different states that, politically, economically, and culturally, more or less differed one from another. On the linguistic scale, the only form of standardization consisted in that the speech of major centers was sometimes considered as a model to reach (Maiden). For quite a long time, Latin remained the language in religious, scientific, administrative, and judicial matters. In literature, however, there was a growing interest in the use of the regional language; the discussion, known as the *Questione della Lingua* ('Language question'), about the nature of a literary language for the Italian Peninsula ended up with the hegemony of the Florentine language, as used in the works of Dante Alighieri (1265–1321), Francesco Petrarca (1304–1374), and Giovanni Boccaccio (1313–1375), and described in Pietro Bembo's grammar *Prose della volgar lingua* (1525).

It was only in 1861, after the movement known as *Risorgimento* ('Resurrection'), that the country was politically unified. At that time, the writer Alessandro Manzoni (1785–1873) had already been confronted with the lack of a national language. For the third version of his famous novel *I promessi sposi* (1840, 'The Betrothed'), he had deliberately chosen the contemporary speech of the cultivated Florentine classes, since literary Florentine missed vocabulary to allow him to write about everyday life. Although Manzoni's novel aimed and obtained national spread, it was rather the literary variant that was diffused in the newly unified country (Maiden), where, around 1861, only a very small part of the literate population had a thorough command of the standard language. Despite the political unification, the linguistic unification took place rather slowly in its beginnings. However, it was boosted by a number of external factors, such as national education, the rise of mass media, general conscription, migration, and urbanization (De Mauro).

Research executed by the Italian opinion investigation institute Doxa in 1991 shows that 12.8% of the country's population claims to use 'dialect' in all or nearly all situations, while 33.6% speaks standard Italian on every occasion. However, most Italians use alternatively their proper 'dialect' and the standard language, depending on the communicative situations they are in (e.g. familiar discussion vs. formal discourse). With regard to this state of affairs, one could think of using the terms bilingualism or diglossia. It seems that, as far as Italy's standard languages and 'dialects' are concerned, the term diglossia is the most appropriate, although Ferguson's definition does not match exactly the Italian situation. Anyway, nowadays, the very large majority of Italy's inhabitants are able to express themselves in standard Italian. Apart from everyday speech, 'dialects' can also be used in literary works and for the major 'dialects', there are translating dictionaries from and to standard Italian.

As for the classification of the different 'dialects', important divisions can be made, but it should be borne in mind that on the geographical plan, the 'dialects' gradually blend into one another and that most distinct borders represent, in fact, gray zones of transition.

Traditionally, three major groups are identified:

—northern 'dialects'
—Tuscan 'dialects'
—central and southern 'dialects'.

It is considered that an imaginary line between the cities of La Spezia and Rimini delimitates the northern zone from the central and southern 'dialects', Tuscan 'dialects' (also situated below the La Spezia-Rimini line) occupying a specific position. This line roughly delimits a number of more or less general features, but it should be clear that these are abstractions and that a large variation exists within each of the 'dialect' groups. In any case, the 'dialects' closest to standard Italian are the Tuscan ones.

In addition to the so-called 'dialects', one can find, in some regions of Italy, other varieties different from standard Italian. These varieties (among which the most important are Sardinian, Friulian, and Ladin) developed from spoken Latin, just like standard Italian and Italian 'dialects'. Their status, however, has been a matter of discussion for quite a long time. Nowadays, most scholars refer to them by using the term 'languages' rather than 'dialects'.

Due to its geographical isolation, the language of the isle of Sardinia is considered as the most conservative of Romance languages, i.e. closest to Latin. Nevertheless, it should be stressed that there is no such thing as a standard Sardinian language: several varieties exist, and they are traditionally classified into four groups (speech of the Logudoro, the Campidano, the Gallura, and of Sassari). As the spread of a national language after 1861 also reached Sardinia, standard Italian is used on most formal occasions. However,

Sardinian varieties are spoken by about 1.2 million people nowadays (Klinkenberg), mostly in informal contexts. Therefore, the term bilingualism (when Sardinian is considered to be a language) and diglossia (when it is regarded as a dialect) is also used in the case of Sardinia.

Friulian and Ladin are two languages spoken in northern Italy. Friulian is spoken in parts of the autonomous region of Friuli-Venezia Giulia. The use of Ladin is limited to a number of valleys in Alto Adige (or South Tirol, part of the autonomous region of Trentino-Alto Adige, see map). Both languages are closely related to Romansh, one of Switzerland's four official languages, and have characteristics in common with northern Italian 'dialects'. Yet, like Romansh, they each represent one of a series of 'dialects'. In the case of Friulian, the variety from Udine and environs is sometimes considered as a base for a standard, but this assertion has also been criticized. Friulian is estimated to be spoken on a regular basis by approximately 430,000 people (Picco). However, it is again standard Italian that is used in formal and official contexts. One might be tempted to use the terms diglossia and bilingualism, but none of them fully accounts for the situation: *diglossia* would imply that Friulian is considered as a 'dialect' of standard Italian and the use of *bilingualism* would mean that both Friulian and Italian are used in all circumstances. Attempts to use Friulian at school and in mass media are not always successful (Francescato 1989); the lack of a Friulian standard appears to be one of the major problems.

As for Ladin, it has, like standard Italian and German, an official status in Trentino-Alto Adige and is probably spoken by about 35,000 people (official figures and estimates). There is a written standard, and Ladin can be taught and used at school. However, the Ladin language is under pressure, as it is less prestigious than standard Italian and German. In fact, these two languages are far more spoken, not only seen in a European context but also within the region itself. Moreover, German and Italian are languages that convey a widespread culture.

Like German, a number of minority languages spoken in Italy have in common that they are an official language in one or more countries or regions outside Italy. Their status within the Italian Republic, however, is not the same for all.

German is an official language in Trentino-Alto Adige, together with standard Italian and Ladin. Schools can be either in German or in Italian, but in both of them the second language is the region's other official language; there are also plurilingual schools with a fifty–fifty division of German and Italian, where Ladin is also taught (Freddi). As far as the linguistic features are concerned, the German variety spoken in Trentino-Alto Adige joins the Tyrolean and Bavarian type of speech (Mazzotta).

French is an official language in the autonomous region of Val d'Aosta, in Italy's most northwest corner. In this region, it also constitutes a compulsory subject in education. Yet another Romance language is spoken by about 90,000 people (Klinkenberg): this language, called Franco-Provençal, represents merely a group of different dialects without a standard. These dialects historically belong to a zone covering a part of France, Switzerland's French-speaking region and Val d'Aosta. In formal and official contexts, however, they are almost completely surpassed by French and Italian.

A similar situation can be found in Piedmont, where small Franco-Provençal, Occitan, and Walser minorities exist. Occitan is also a Romance language, mainly spoken in southern France, yet remaining a minority language there as well. Occitan has to cope with problems related to standardization, but like Franco-Provençal it is recognized as a minority language by national law and is promoted by various cultural associations. The same holds for the speech of the Walser, which is a variety of German spoken by a small group of people in Val d'Aosta and Piedmont.

Catalan, the official language of the region of Catalonia in Spain, is spoken by a minority of 15,000–20,000 people in the Sardinian town of Alghero, where it is also a subject in education (Klinkenberg).

Slovenian is spoken by approximately 20,000–30,000 people in Friuli-Venezia Giulia, close to the Slovenian border; there are a number of schools that provide education in Slovenian (Francescato-Ivašič Kodrič).

Serbo-Croatian is estimated to be spoken by a small comunity of 4,000–5,000 people in central Italy (Mazzotta).

Albanian and Modern Greek are spoken by scattered communities in southern Italy; it is mostly elderly people who seem to hold on to their language.

In a country where so many language varieties are used, different phenomena due to language contact can be expected. In regions where one or more languages coexist with standard Italian, these languages (French, German, etc.) may influence standard Italian. However, it is clear that the position of standard Italian as the Republic's official language remains first, as its knowledge is boosted through public education, mass media, etc.

As for the contact between standard Italian and Italian 'dialects', it is obvious that the spoken and—to a lesser extent—written use of standard Italian bears marks of 'dialectal' features, such as differences in pronunciation or in vocabulary. An example of the former is the Roman -NN- pronunciation, where standard

Italian has -ND- (*annamo* for *andiamo* 'we are going'); an example of the latter can be the use of the Venetian word *campo* for *piazza* 'square'. By the way, 'dialectal' words can be accepted in standard Italian, like the word *tortellini* ('tortellini', i.e. a type of pasta), which has its origin in Emilia (Zolli).

But the standard language also leaves its traces in 'dialectal' speech. This mainly happens in the 'dialects's' sound system and vocabulary (where a local term is abandoned in favor of an Italian one; e.g. Piedmontese ['pejla] 'pan' is replaced by [pa'dela], from standard Italian *padella*) (Sobrero 1997). The diffusion of the standard language also influences the global use of the 'dialect'. Referring to the contemporary situation, it is possible to speak of an abandonment of 'dialect', which is, however, far less frequent in rural areas than in urban zones, where several people from different origins live together and where various cultures coexist.

Recently, the term *italiano popolare* ('popular Italian') has been used to indicate a kind of substandard Italian. This substandard Italian originates from an imperfect acquisition of the standard language, possibly—but not necessarily—due to a 'dialectal' background. Some typical features of popular Italian are:

English	Standard Italian	Popular Italian
it seems to me ...	*A me* pare ... (or *Mi* pare ...)	*A me mi* pare ...
to have	*avere*	*averci* (= *avere* + *ci* 'there')
the place where we have been	il posto *dove* siamo stati	il posto *che* siamo stati ('the place *that* we have been')

Popular Italian is not a language on its own, but a whole of characteristics that differ from standard Italian. These characteristics can be widespread through the peninsula (like the examples above), but sometimes they are limited to one or more (parts of) regions.

The campaign that aimed, shortly after Italy's unification, to promote 'standard Italian' as a national language, seems to have reached its goal. Local and regional types of speech often remain lively, which leads to intense contact between the different codes.

References

De Mauro, Tullio. 1963. Storia linguistica dell'Italia unita. Bari: Laterza. 2nd edition, 1970.

Ferguson, C.A. 1959. Diglossia. Word 15. 325–40.

Francescato, Giuseppe. 1989. Friaulisch: Soziolinguistik. Sociolinguistica. Lexikon der Romanistischen Linguistik (LRL). Band/Volume III. Die einzelnen romanischen Sprachen und Sprachgebiete von der Renaissance bis zur Gegenwart. Rumänisch, Dalmatisch/Istroromanisch, Friaulisch, Ladinisch, Bündnerromanisch. Les différentes langues romanes et leurs régions d'implantation de la Renaissance à nos jours. Le roumain, Dalmatico/Istroromanzo, Friulano, Ladino, Le romanche, ed. by Günter Holtus, Michael Metzeltin, and Christian Schmitt. Tübingen: Niemeyer.

Francescato, Giuseppe, and M. Ivašič Kodrič. 1983. La comunità slovena in Italia. L'Italia plurilingue, ed. by Giovanni Freddi. Bergamo: Minerva Italica.

Freddi, Giovanni. 1983. Maggioranze, minoranze e plurilinguismo nella Provincia di Bolzano (Südtirol-Alto Adige). L'Italia plurilingue, ed. by Giovanni Freddi. Bergamo: Minerva Italica.

Klinkenberg, Jean-Marie. 1994. Des langues romanes. Louvain-la-Neuve: Duculot.

Lepschy, Anna Laura, and Giulio Lepschy. 1977. The Italian language today. London: Hutchinson.

Maiden, Martin. 1995. A linguistic history of Italian. London and New York: Longman.

Maiden, Martin, and Mair Parry (eds.) 1977. The dialects of Italy. London and New York: Routledge.

Mazzotta, Giuseppe. 1983. Profilo linguistico dell'Italia odierna. L'Italia plurilingue, ed. by Giovanni Freddi. Bergamo: Minerva Italica.

Picco, Linda. 2001. Ricercje su la condizion sociolenghistiche dal furlan. Ricerca sulla condizione sociolinguistica del friulano. Udine: Forum.

Rohlfs, Gerhard. 1949–1954. Historische Grammatik der Italienischen Sprache und ihrer Mundarten. I Lautlehre, II Formenlehre und Syntax, III Syntax und Wortbildung. Bern: Francke.

Sobrero, Alberto. 1997. Italianization of the dialects. The dialects of Italy, ed. by Martin Maiden and Mair Parry. London and New York: Routledge.

Zolli, Paolo. 1986. Le parole dialettali. Milano: Rizzoli.

The data concerning Italy's area and population are the figures provided for 2000 by the Italian institute ISTAT (www.istat.it).

The data from the opinion investigation institute Doxa can be obtained through www.doxa.it.

ALEXANDER LOENGAROV

See also **Bilingualism; Diglossia; Indo-European 4: Romance; Italian; Switzerland**

J

Jakobson, Roman

Roman Jakobson was one of the major linguists, literary theorists, and semioticians of the twentieth century, and a leading proponent of functional structuralism.

Structural Phonology, Child Language, and Aphasia

Jakobson helped to found the Prague Linguistic Circle, the cradle of (functional) structuralism, in 1926. For the Prague School, language serves for communication and thus its internal structure has to be studied from the standpoint of the tasks it performs. Jakobson and Nikolai Trubetzkoy argued that phonological systems are structural wholes, based on their relational properties, and that language change has to be seen in terms of systems. They investigated types of phonological systems and established the study of linguistic affinities (*Sprachbünde*)—how languages that are geographically and culturally related may acquire traits in common.

Jakobson also turned his attention to the acquisition of phonology by children and its dissolution in aphasics and demonstrated that the order of acquisition goes from universal to nonuniversal, simple (optimal) structures to more complex ones. He also showed that aphasic loss follows a reverse order.

Theory of Distinctive Features and Functional View of Sound

Jakobson showed that phonemes such as /b/ and /m/ in *bat* and *mat* are not the smallest constituents of language. Rather, distinctive features are. For example,

/b/ in *bat* is differentiated from /m/ in *mat* as oral vs. nasal. He defined a small set of such features that underlie the phonological systems of all the languages of the world. He also showed the relevance of these notions for grammatical meaning, which he decomposed into semantic features.

Invariance and Relational Structure

Invariance was, for Jakobson, the dominant topic and methodological device underlying his research: any sign is defined by its invariant properties. Across languages as well, he also insisted on the importance of universal invariants (language universals), establishing the equivalence of diverse signs. Invariance for Jakobson was always invariance in relationship (equivalence), and the primary type of relation in language is that of binary opposition, phonological (e.g. nasal vs. oral), grammatical (e.g. plural vs. singular), or lexical (e.g. *near* vs. *far*), etc. Opposition is based on an asymmetry in the relationship between a marked (focused or weighted) and an unmarked (neutral) term: e.g. unmarked oral, singular, *far* vs. marked nasal, plural, *near*. For Jakobson, markedness is a fundamental way by which we organize our linguistic, semiotic, and sociocultural systems in general.

Functions of Language

For Jakobson, the *raison d'être* for language is communication and language is suited to various communicative goals, which in turn are correlated with the speech event in which language is used. He defined

six primordial functions in terms of a focus on one of the facets of the speech event: (1) emotive function (focus on the speaker)—e.g. intonation showing anger; (2) conative function (focus on the addressee)—e.g. imperatives and vocatives; (3) referential function (focus on the context)—e.g. talk about the real world; (4) poetic function (focus on the message)—e.g. poetry; (5) metalingual function (focus on the code)—e.g. definitions of words; and (6) phatic function (focus on the contact)—e.g. 'hello, do you hear me?' These functions may be dominant in a message or may be subsidiary: e.g. a referential message may also carry expressive information about the opinion of the speaker.

Metaphor and Metonymy: Similarity/Contiguity

Jakobson also analyzed the relation between communicative processes and properties of linguistic structure. First, he distinguished the two fundamental operations: selection (substitution) for encoding (production) and combination for decoding (comprehension). Then, he contrasted two types of relations: similarity (all types of equivalence) and contiguity (temporal and spatial neighborhood). Put simply, the elements in a selection set are normally associated by similarity, and those in combination by contiguity. For Jakobson, similarity/contiguity is a fundamental polarity of language, texts, culture, and human thought (cognition) in general. He proposed using it as a way of categorizing types of aphasic disturbances and also of defining the poetic function: in poetry, where focus on the message is dominant, equivalence (similarity) relations help to build the combination. Thus, rhyme, alliteration, parallelisms, metaphor, etc. help to structure the poetic text, whereas in prose contiguity is the essential constructional principle.

Grammar and Semantics

Jakobson also focused on the function of grammatical categories. For example, he worked on those elements whose general (invariant) meaning in the code can only be specified by taking into account their use in speech events, e.g. deictic pronouns designate speaker (*I*) and addressee (*you*). Thus, language encodes pragmatic factors of the context of utterance.

Grammatical categories (both morphological and syntactic), for Jakobson, are obligatory, whereas particular lexical categories (e.g. words referring to space) are optional. Thus, he provided a semantic and operational approach to the relation between language and cognition: grammatical categorizations provide the necessary patterns of thought.

Jakobson's Legacy in Linguistics

Many of Jakobson's contributions to twentieth-century thought have become a permanent part of American and European views of language. He contributed to linguistics such concepts as (distinctive) feature, binary opposition, structuralism, markedness, and universals. He also showed the importance for linguistics of child language acquisition, aphasia, poetry, the act of communication (language usage), the meaning of grammatical categories, and the systematicity of language change. He has had such a towering role in linguistics that his work has defined the field itself and many of his concepts and discoveries are now thought to be commonplace or self-evident.

Jakobson was a leading proponent of structuralism, arguably one of the most influential trends not only in linguistics but in other humanistic areas as well. His functional viewpoint has inspired functionalist approaches more generally and abetted developments such as discourse analysis, text linguistics, ethnography of communication, anthropological linguistics, sociolinguistics. Work in pragmatics has led linguists to explore the boundaries of their discipline with the neighboring fields of anthropology, sociology, mythology, philology, and philosophy in a truly Jakobsonian interdisciplinary spirit.

Jakobson stressed the need to search for general laws governing all linguistic systems and thus helped to foster the current interest in universal and law-governed, general (explanatory) properties of language. His seminal work on child language and aphasia has been signaled as the point of departure of a new era in linguistics because it showed the relation between child language and phonology, and it launched psycho- and neurolinguistics and contributed to contemporary work on the functional organization of the human brain.

The widespread movement that presently investigates universal properties of language, both universal grammar (Noam Chomsky) and language universals (Joseph Greenberg), is deeply indebted to Jakobson. Many scholars agree with Jakobson that typology is important for the study of language change, areal linguistics, and historical reconstruction. His seminal idea that changes must always be treated in view of the system that undergoes them is the essential premise for textbooks in historical linguistics.

Many linguists have claimed that the distinctive feature was Jakobson's greatest insight and, after the phoneme, the most significant step forward in modern phonology. Generative phonology recognizes Jakobson as a founding father, markedness is now widely used in linguistics and other disciplines, and componential analysis—analyzing an item into smaller properties

and representing them as a combination—has been adopted by many.

With the breadth and depth of his knowledge, his brilliant imagination, his prodigious output, the originality of his accomplishments, and his international influence on a variety of fields, Roman Jakobson has been recognized worldwide as one of the major creative minds of our century.

Biography

Roman Jakobson was born in Moscow, Russia on October 11, 1896. He was educated at Moscow University, in the Slavic section. He was awarded the Buslaev Prize in 1916 for his work on North Russian folk epics, and the master's was awarded in 1918. He helped to found the two groups known now as Russian Formalism in 1915–1916. He moved to Prague in 1920. In 1926, he was cofounder of the Prague Linguistic Circle and was its vice-president until 1939. He received his doctorate in 1930 from Prague University. In 1933, he began teaching at the T.G. Masaryk University at Brno, and assumed the chair of Russian Philology and Old Czech literature at Brno in 1937. In 1939, he fled Nazi invasion and went to Scandinavia. He was visiting lecturer in Copenhagen and Oslo until 1940, and then in Uppsala until 1941. In 1941, he took a freighter to the United States, and taught at the Ecole Libre des Hautes Etudes in New York in 1942–1946. In 1943, he was cofounder of the Linguistic Circle of New York and its vice-president until 1949. From 1943 to 1946, he was visiting professor of general linguistics at Columbia University and became T.G. Masaryk Professor of Czechoslovak Studies in 1946. In 1949, he was named Samuel Hazzard Cross Professor of Slavic Languages and Literatures at Harvard University, and in 1960 he became professor of general linguistics as well. In 1957, he was also named Institute Professor at MIT. He was President of the Linguistic Society of America in 1956, and received the International Prize for Philology and Linguistics in 1980 and the Hegel Prize in 1982. Jakobson died in Cambridge, Massachusettes on July 18, 1982.

References

A Tribute to Roman Jakobson 1896–1982. New York: Mouton. 1983.

Armstrong, J. Daniel, and C.H. van Schooneveld (eds.) 1977. Roman Jakobson: echoes of his scholarship. Lisse: Peter de Ridder Press.

Bradford, Richard. 1994. Roman Jakobson: life, language, art. London: Routledge.

Holenstein, Elmar. 1976. Roman Jakobson's approach to language. Bloomington: Indiana University Press.

Jakobson, Roman. 1990. A complete bibliography of his writings, 1912–1982, ed. by Stephen Rudy. Berlin: Mouton.

Jakobson, Roman, and Krystyna Pomorska. 1980. Dialogues. Cambridge, MA: MIT Press.

_____. 1987. Language in literature, ed. by Krystyna Pomorska and Stephen Rudy. Cambridge, MA: Harvard University Press.

_____. 1990. On language, ed. by Linda Waugh and Monique Monville-Burston. Cambridge, MA: Harvard University Press.

_____. 1962. Selected writings I, phonological studies. The Hague: Mouton; 2nd expanded edition, 1971; II, Word and language, The Hague: Mouton, 1971; VII, Contributions to comparative mythology; studies in linguistics and philology, 1972–1982, Berlin: Mouton, 1985; VIII, Major works, 1972–1982, Berlin: Mouton, 1987.

Waugh, Linda R., and Monique Monville-Burston. 1990. Introduction: the life, work and influence of Roman Jakobson. In Jakobson.

LINDA R. WAUGH

See also **Phonology; Trubetzkoy, Nikolai Sergeyevich**

Japanese

Japanese is the national and official language of Japan, spoken by approximately 126 million people on the Japanese islands, as well as by Japanese minorities in Hawaii, North and South America, and other parts of the world. The language is called *Kokugo*, 'language of the country' in Japan, and *Nihongo*, literally 'Japan-language', when spoken by nonnative speakers. In terms of number of speakers, Japanese ranks eighth on the list of major languages of the world.

Japanese is written with ideographic characters, called *kanji*, originally borrowed from China, and also with two syllabaries, the *hiragana* and *katakana*, developed in Japan by simplifying the Kanji Chinese characters. A *kanji* is connected to semantic content and has several different pronunciations, while *hiragana/katakana* represent syllables. Roughly, *kanji* are used for content words, e.g. 'sun' or 'read', *hiragana* for grammatical endings and grammatical function

words, e.g. *-ta* past tense or *moshi* 'if', and *katakana* for loan words, e.g. *erebeetaa* 'elevator'. Additionally, Roman script is used for abbreviations, e.g. NHK, and words cited from foreign languages. Traditionally, Japanese is written vertically and from right to left, but today horizontal left to right writing is also used, depending on text type and space.

The question of any language family membership of Japanese is still disputed. Japanese is not related to Ainu, formerly spoken on Japan's biggest northernmost island Hokkaido, the Kurile islands, and Sakhalin (an island off the northeast coast of the Asiatic continent). Nor does Japanese belong to the Sino-Tibetan language family spoken on the Asiatic continent, the most prominent representative of which is Chinese. The most widely discussed hypothesis today is an affiliation to Korean (spoken in Korea) and to Altaic languages (Turkic, Mongolian, Manchu, and Tungusic languages, spoken in continental Asia).

Japanese shares a substantial number of grammatical characteristics with both Korean and Altaic languages: word order in the sentence; suffixal inflection (with endings rather than with prefixed elements); postpositions (functioning like English prepositions, but following the noun); the existence of 'converbs' (verb forms that are used for clause linking instead of conjunctions); question formation via particles at the end of a sentence; vowel harmony (a restriction on vowels occurring together in a word, of which there are traces in Old Japanese); and the absence of a gender system, articles, and relative pronouns. Attempts to establish lexeme correspondences between Altaic languages and Japanese, and to reconstruct a common proto- (ancestral) language, however, have not led to a significant amount of convincing phonological correspondences. The absence of complex consonant clusters (e.g. tk, pl, str), and the near absence of syllable-final consonants in the Japanese sound pattern, contrasts with Altaic languages, where these uses do occur. Additionally, vowel harmony occurred only briefly in the linguistic history of Japanese, and inflectional endings on verbs for persons involved in the event being discussed are absent in Japanese.

The sound pattern, the open (vowel-final) syllables, and the basic bisyllable pattern of words in Japanese are reminiscent of a neighboring language family, the Austronesian languages (spoken in Taiwan, the Philippines, Indonesia, Oceania, and Madagascar), in which these same properties are characteristic. But again, a sufficient number of phonological correspondences between Japanese and Austronesian have not been established, and sentence structure is disparate. It is therefore assumed by some scholars that Japanese might be a hybrid language, amalgamating features of Altaic and Austronesian languages. Besides this, a

relation to Dravidian languages (spoken in the south of India) is hypothesized, as Dravidian languages also share the grammatical features listed above for Altaic languages.

The earliest Japanese writings date back to the eighth century CE. The language then developed from Old Japanese to Modern Japanese between the twelfth and sixteenth centuries. Late Old Japanese and Middle Japanese are considered today as the literary language (*bungo*). The term 'Classical Japanese' is used either to correspond to *bungo*, or may designate any stage of the language between the eighth and fourteenth centuries. This epoch produced a rich literature of lasting importance by court nobles, especially court ladies, both in prose and verse.

After Japan was politically unified in the seventh and eighth centuries, the greater Kyoto region in the west of the main island of Honshu became the political and cultural center in the eighth century, which it remained, almost without a break, for about a thousand years. The political center was then moved eastward to Tokyo at the beginning of the seventeenth century, and gained cultural authority until the end of the eighteenth century. Today, the speech of the Tokyo region, adorned by influences from the Kyoto variety and interspersed with traits of the prevailing local dialect, especially accentual features, functions as the standard national and official language.

Japanese is rich in dialectal variation. Several different classifications have been proposed, but two major dialect divisions are generally recognized: (1) between the Eastern dialect group spoken in the northeastern half of the main island of Honshu up to the northern island of Hokkaido and the Western dialect group spoken in the southwestern part of Honshu, and (2) between the Western dialects and the dialects spoken on the island of Kyushu, south of Honshu. The East–West dialect boundary runs through Central Honshu between the Niigata and Toyama prefectures in the north, along the Japanese Alps, and between the Shizuoka and Aichi prefectures in the south. The boundary is constituted by the existence and pronunciation of several sounds, by the accent pattern, and by a number of inflectional forms of the verb. Other dialects are differentiated by their sound and accent patterns, lexicon, and functionality of case markers.

A fourth linguistic area is formed by the islands of Okinawa, south of Kyushu. The language of Okinawa is sometimes considered to be another major dialect of Japanese, but it is classified by other scholars as a separate language, known as Ryukyuan or Luchuan. The latter approach is supported by the mutual unintelligibility of Japanese and Ryukyuan. Ryukyuan is estimated to have split from the Japanese stock around the beginning of the Christian era and was the language of

an independent kingdom between the fifteenth and seventeenth centuries CE, before the Ryukyu islands were incorporated into Japan. Ryukyu is divided into three larger dialects, which are mutually unintelligible both among themselves and with Japanese dialects. While Ryukyuan has a different sound pattern with fewer vowels than Japanese, sound correspondences were easily established. Morphology (word structure) and syntax (clause structure) are quite similar to Japanese, but Ryukyuan preserves the lexical, morphological, and syntactic features found in Old Japanese, having also introduced inflectional forms:

	Japanese (Tokyo)	**Ryukyuan (Naha)**
'nose'	*hana*	*hana*
'wind'	*kaze*	*kazi*
'hand'	*te*	*ti'i*
'smoke'	*kemuri*	*kibusi*
'bird'	*tori*	*tu'i*

Among the 1.3 million inhabitants of the Ryukyu islands, Ryukyuan is spoken mainly by the older generations. Those under 20 years of age speak only Japanese. Thus, Ryukyuan is a severely endangered language. Although some private broadcasting services provide news in Ryukyuan, its prospects for surviving are bad for political reasons.

Japanese grammar and lexicon present a number of phenomena that are of theoretical interest to general linguistics. Japanese has a pitch-accent pattern, which differs from both stress-accent patterns (as in English)—in which accent is marked by sound pressure or loudness—and tone languages such as Chinese—in which each syllable bears one fixed pitch out of a pool of four pitch types. In Japanese, every word has a pitch 'contour' with, at most, one pitch drop. The place of the drop from high to low pitch in a word distinguishes meaning: e.g. *hashi-ga* with the pitch pattern high–low–low (i.e. high pitch on *ha*) means 'chopsticks', while *hashi-ga* with the pitch pattern low–high–low (i.e. high pitch on *shi*) means 'bridge', and *hashi-ga* with the pitch pattern low–high–high (i.e. high pitch on *shi* and *ga*, no drop) means 'corner'.

A complex system of honorifics (expressions of respect/politeness) plays an important role in any Japanese utterance. Two kinds of honorifics are interwoven, namely, the expression of politeness toward the interlocutor, and the expression of respect toward a person one is talking about. Appropriate use of honorifics is determined by the social context of the utterance, and is socially obligatory. Politeness to the interlocutor is expressed by an inflectional ending on the verb (*-masu*); respect toward a person being mentioned is shown by the choice of special verbs or derivational forms (word formation products) of verbs and nouns. There is a neutral form for every verb (used with reference to equally ranked persons); an elevating form, which is used with reference to a socially superior person; and a humble form, which is used with reference to the speaker or the speaker's group:

	Neutral	**Elevating** 'a respected person does X'	**Humble** '(speaker) does X for/to a respected person'
'read'	*yomu*	*o-yomi-ni naru*	*o-yomi-suru* '...
'exist'	*iru*	*irassharu*	*oru*
'come'	*kuru*	*irassharu*	*mairu*
'do'	*suru*	*nasaru*	*itasu*
'eat'	*taberu*	*meshiagaru*	*itadaku*

Respect forms also exist for a number of nouns, especially kinship terms. The prefixes *o-* and *go-* are honorific:

Own thing/kin		**Respected person's thing/kin, address term**
tegami	'my letter'	*o-tegami* 'your/his/her/their letter'
hon	'my book'	*go-hon* 'your/his/her/their book'
haha	'my mother'	*o-kaasan* 'your/his/her/their mother', 'mother!'
chichi	'my father'	*o-toosan* 'your/his/her/their father', 'father!'
ane	'my elder sister'	*o-neesan* 'your/his/her/their elder sister', 'sister!'
otooto	'my younger brother'	*o-toosan* 'your/his/her/their younger brother'

An elaborate honorific system is also present in Korean.

As in Altaic languages and Korean, Japanese is rich in onomatopoetics (sound-symbolic words), which are conventionalized mimetic expressions of natural sounds, mental and emotional states, sensations, and physical states and manners. These words are usually used as adverbs. Examples are *wan-wan* (bow-wow), *patan* (with a bang), *zaa-zaa* (sound of downpour), *kossori* (stealthily), *ira-ira* (nervously), *zara-zara* (rough texture), *hara-hara* (falling down softly; fearfully), and *bara-bara* (falling of heavy objects).

Most Japanese loanwords are borrowings from Chinese and have a status comparable to Latinate words in European languages. Other loanwords come from European and other Asian languages, especially from English. Loan-word adaptation both to the Japanese sound pattern and word classes resulted in compromise forms, which introduced new sounds and syllable-final consonants into Japanese. Native words and loanwords with similar meanings may coexist and usually have specialized meanings: e.g. *torikeshi* (native), *kaiyaku* (Chinese loan), and *kyanseru* (from

English 'cancel') all mean 'cancellation', but the first example has a very general meaning, while the Chinese loan is used for formal occasions such as in contracts and the English loan for reservations or appointments.

A phenomenon of Japanese word formation of great theoretical importance occurs when from a clause construction larger than one word, the last element may undergo word formation without any other changes in the construction taking place. For example, the phrase *sake-o nomita-sa ni* (out of desire of drinking sake) is formed from the clause *sake-o nomitai* (s.o. wants to drink *(nomitai)* sake) with the help of the ending *-sa.* This is remarkable in view of the theoretical assumption that word formation is a word-level process and should not be accessible beyond the word level.

On the morphological (word structure) level, Japanese is agglutinative, i.e. various grammatical endings are chained together to form a word. Japanese has word classes of nouns, verbs, and adjectives, although the latter are a subclass not of nouns but of verbs, and they are inflected in the same way as verbs.

The function that a noun fills in a clause is signaled by elements loosely attached to the end of a noun, which are traditionally called 'postpositions'. A postposition is a word that fills the function of a preposition ('at', 'in', 'for', etc.), but follows the noun instead of preceding it. In Japanese, postpositions fill the functions of case: e.g. *Hanako-ga* 'Hanako' (nominative), *Hanako-no* 'Hanako's' (genitive), *Hanako-o* 'Hanako' (accusative), *Hanako-ni* 'to Hanako', *Hanako-to* 'with Hanako', etc. Postpositions are also characteristic of Altaic languages.

The persons involved in an action are not usually explicitly expressed, neither on the verb nor by nouns accompanying it. The bare verb constitutes a minimal sentence, and the persons involved must be deduced from the larger context: e.g. *yomu* is a complete clause that may mean, 'I read it/something', as well as 'you/he/she/it/we/you/they read it/something'. The persons involved in the action can be understood from a range of verb pairs sharing a common root (basic component). One verb of a pair signals that s.o./sth. other than the speaker or her group is acting (so-called 'intransitive'), while the other signals that the speaker is acting on s.o./sth. (so-called 'transitive'):

Nonspeaker acts		Speaker acts	
agaru	'rise'	*ageru*	'raise'
tomaru	'stop'	*tomeru*	'stop'
aku	'open'	*akeru*	'open'
okureru	'be late'	*okurasu*	'postpone'
naku	'cry'	*nakasu*	'make cry'
sakeru	'split'	*saku*	'split'

Furthermore, Japanese verbs are classified according to their behavior in tense and aspect (presentation of an event as 'bounded', having a relevant beginning and/or ending, or 'unbounded', no relevant beginning, or ending). The *-te iru* form (a kind of continuous form) of a great number of verbs refers to a continuing/continuous event, e.g. *yomu* (read), *miru* (look), *furu* (fall), and *kangaeru* (consider). For example, *yonde iru* translates as, 's.o. is reading'. With other verbs such as *shinu* (die), *sameru* (wake up), and *mitsukaru* (be found), the *-te iru* form expresses a result: *shinde iru* means, 's.o. is dead'. A third class lacks this form entirely, while for still other verbs, it is the only form that can be used to refer to present tense.

Japanese has several inflectional verb forms that cannot be used as the predicate form, i.e. the kernel, of an independent sentence, but are the predicate of a dependent clause or in connection with an auxiliary verb, such as the form ending in *-te*; these are used for clause linking. Such verb forms are known as 'converbs', which are also typical of Altaic languages.

An interesting grammatical form of a verb is the so-called passive, formed with the ending *-(r)are-*. This form is used for events that are not controlled by the person to whom they are happening, e.g. *shinobareru* (something spontaneously comes to my mind) or *ame-ni furareta*, literally 'I was fallen (on) (*furareta*) by the rain (*ame-ni*)', meaning 'I was adversely affected by the rain falling'. The 'passive' is also used for the expression of ability, such as *nerarenai* (I cannot sleep), and as an honorific form, such as *korareru* (a respected person comes).

Japanese is also a classifier language, i.e. it makes use of special elements in counting. Nouns in Japanese cannot be counted directly, as in English: 'one book', 'two books', etc. Instead, the numeral element is combined with another element that expresses the properties of the noun being counted (such as shape, kind, or function) or measures it. This classifier is then used in conjunction with the noun. Thus, the element *-ko* is used for small round objects, *-hon* for cylindrical objects, *-ri* and *-nin* for persons, *-dai* for machines, *-ken* for houses, *-sai* for age, *-hai* for cupfuls, and so forth. A counting construction takes the form *hon issatsu* (one book, literally 'book one-bound entity'), *hon nisatsu* (two books), and so forth, with *issatsu* and *nisatsu* being the classifiers for bound objects.

On the clause level, the Japanese word order is subject–object–predicate, as in *Hondasan-wa Fujiisan-o mita* (Mr./Ms. Honda saw Mr./Ms. Fujii). Modifying elements precede the word they modify, according to the patterns *kono ie* (this house), *akai ie* (red house), *watashi-no ie* (my house), *Hondasan-no ie* (Mr./Ms. Honda's house), and *Hondasan-ga mita ie* (Mr./Ms. Honda saw (*mita*) house (*ie*)), which is the equivalent of the English, 'the house that Mr./Ms. Honda saw'. Relative pronouns (such as the English 'which') do not exist. Adverbs likewise precede verbs, according to the

pattern *yukkuri-to aruku* (s.o. walks slowly, or literally slowly walks). A complicated topic discussed in the field of Japanese clause structure is configurationality, or whether or not the elements of a clause are hierarchically structured. As there are arguments both for and against Japanese being configurational, the question remains open.

An aspect of clause structure for which Japanese has become famous is the existence of two structural components of the sentence ('phrases') that are candidates for serving as the subject. These components are identifiable by the markers at their ends, *-wa* and *-ga*, respectively. They have different conditions of usage and varying functions, and they may occur simultaneously in one sentence (known as the 'double subject construction'), as in *zoo-wa hana-ga nagai* (the elephant has a long trunk), or literally, 'the elephant (*zoo-wa*, topic), the nose (*hana-ga*, subject) is long (*nagai*, predicate)'. The component marked with *-ga* is the grammatical partner of the predicate and therefore the subject; the component marked with *-wa* is the topic, i.e. the thing or person about which the speaker says something. The topic may have a looser relation to the predicate, as in *boku-wa unagi da* (eel for me), or literally 'I (*boku-wa*) it is (*da*) eel (*unagi*)'. The topic construction can be better understood if translated by the English 'as for . . . '; for example, 'as for the elephant, its trunk is long'. Japanese shares the existence of such a topic construction with Korean.

Japanese discourse, or larger bodies of spoken or written text, is thoroughly studied as well. Topics of interest in narrative discourse are the organization of a text and the elements used for signaling its structure; the way the persons involved in a story are handled linguistically to maintain a perspective; and the formal chaining of events. In the study of conversation, strategies used for turn-taking, sustaining the audience's attention, and confirming attention are studied, as well as verbal signs used in such interaction, such as absence of grammatical elements, reordering of components of a clause, sentence-final particles (small words), echoing, and affirmative sounds and gestures.

References

Hinds, John. 1986. Japanese. London: Croom Helm.
Jacobsen, Wesley M. 1992. The transitive structure of events in Japanese. Tokyo: Kuroshio.
Kishitani, Shoko. 1985. Die Person in der Satzaussage (Person in sentential predication). Wiesbaden: Harrassowitz.
Kuno, Susumu. 1973. The structure of the Japanese language. Cambridge: MIT Press.
Kuroda, Shige-Yoshi. 1972. The categorical and the thetic judgment: evidence from Japanese. Foundations of Language 9. 153–85.
Lewin, Bruno. 1959, 1996. Abriß der japanischen Grammatik auf der Grundlage der klassischen Schriftsprache (An outline of Japanese grammar based on the classical written language). Wiesbaden: Harrassowitz.
Martin, Samuel E. 1975. A reference grammar of Japanese. New Haven: Yale University; Rutland: Tuttle, 1988.
Maynard, Senko K. 1998. Principles of Japanese discourse. A handbook. Cambridge: Cambridge University Press.
Miller, Roy Andrew. 1967. The Japanese language. Chicago: University of Chicago Press.
Sansom, George. 1928, 1968. An historical grammar of Japanese. Oxford: Clarendon Press.
Shibatani, Masayoshi (ed.) 1976. Japanese generative grammar (Syntax & Semantics 5). New York: Academic Press.
———. 1990. The languages of Japan. Cambridge: Cambridge University Press.

JOHANNA MATTISSEN

See also **Ainu; Altaic; Chinese and Japanese Traditional Grammar; Korean**

Japanese Sign Language

Japanese sign language (JSL, or NS after its Japanese name, *Nihon-Shuwa* (alternatively, *Nihon Syuwa*), literally 'Japanese hand-talk') is the dominant language of the deaf community of Japan, and to a lesser extent of the hard-of-hearing community. Owing to the establishment of deaf schools during the time of the military occupations in World War II, JSL is also cognate with the (South) Korean and Taiwanese sign languages. In the latter case, the dialects of Taipei and Tainan are said to derive from the Osaka and Tokyo dialects of JSL, respectively (see Smith 1990). JSL shares an extremely high degree of basic lexical cognates with each of the other two members of the Japanese sign language family.

Sociolinguistics

The Tokyo dialect is the socially dominant variety of Japanese sign language, owing in large part to efforts of the Japan Association of the Deaf and broadcasts of sign

language courses and sign language news on the NHK (Nihon-Housou-Kyoukai = Japanese Broadcasting Corporation, or public television) educational channel. Due to these influences, differences between dialects, although present, are declining. Lexical differences continue to be considerable in some dialects, particularly among elderly signers. Although little research has been undertaken in this area, syntactic differences also arise, affecting auxiliary verb and subject/object–verb agreement in western Japan.

As is the case with perhaps all minority languages, JSL is more or less influenced by the majority language—in this case, by spoken (and written) Japanese. The extent of influence varies according to several factors. First, JSL varies with the education of the speaker (and whether the speaker obtained his or her education in the period between the world wars and up to the early 1990s, when the Ministry of Education (Monbushou) took over control of deaf school administration and promoted a strictly oral education). Considerable difference exists, for example, between the signing of older and younger people. Significant numbers of older signers use little or no finger-spelling, and younger signers tend to allow mouthing patterns to play a much greater role in discourse. Second, JSL also varies with the speech situation (more formal situations often result in signing closer to the signed Japanese than to JSL per se), and on the judged signing proficiency (and hearing status) of the interlocutor. It would be better, therefore, to speak of a continuum from Japanese sign language to signed Japanese. Signed Japanese uses mostly JSL signs, but with Japanese word order and without most of the grammatical features of JSL (and, it should be noted, without many of the grammatical features of spoken Japanese either). Two notable (and highly noticeable) features of signed Japanese (as opposed to JSL) are the extensive use of the sign for *exist* (the inanimate form) as a copular (or linking) verb (Japanese *desu*) and the ubiquitous use of a question particle sign at the end of all questions (equivalent to Japanese *ka*); both signs have an extremely limited use with these meanings in native signing (the latter marking perhaps only a focalized question).

The differences between JSL and signed Japanese, and the limited comprehensibility of the latter to many deaf people, has a deleterious impact, especially in the fields of education and interpreting. Since almost all deaf school teachers can hear, with spoken Japanese as their primary language, and since few of those who sign have contact with JSL-signing deaf adults, the signing produced in deaf schools is almost totally signed Japanese. Also, in the field of interpreting, much interpreting is performed by volunteers, with little linguistic training or real understanding of the dif-

ferences between the two sign systems; hence, signing here is also often on the Signed Japanese end of the spectrum. In both deaf schools and interpreting situations, the reduced degree of comprehensibility of the sign system used means that much of the content is understood only partially.

Visual Cue Structure

As proposed for other sign languages such as American sign language (ASL), the structural features of JSL consist of (1) hand shape, (2) handedness (one vs. two-handed signs), (3) location of the sign in signing space (including relative to a passive base hand), (4) motion (both the hand and arm as a whole, and fingers and parts of the hand to affect hand-shape change), (5) palm and finger-tip orientation, and (6) nonmanual signals. Of all the features entering into JSL signs, the most significant studies have concentrated on hand-shape structure and nonmanual signals.

JSL comprises approximately 50 hand shapes, the vast majority of which are presumably phonemically distinctive. Of these phonemic hand shapes, perhaps half a dozen or so form the basic building blocks of the language. JSL lacks some hand shapes that are present in other sign languages (e.g. ASL /t/), and conversely possesses some that are absent elsewhere (the JSL hand shape with an extended middle finger is the prime example). In each of these cases, the given hand shape is taboo in the other (dominant as well as deaf) culture.

In contrast to the half dozen basic hand shapes, which are widespread and carry a high functional load, some hand shapes have exceedingly limited distribution, at times restricted to a single sign and a single meaning, giving them marginal status. Thus, for instance, the hand shape with all but ring finger extended (and ring finger totally bent) occurs only in the sign for *swallow* (bird). Virtually the same hand shape, but with only the proximal joint of the ring finger bent, occurs only in the sign for *medicine*. The hand shape with only the little finger bent is limited to signs with the meaning 'eight' (also '800' and '8,000'), and the same hand shape with all the extended fingers bent is limited to the meaning '80'.

Quite a few JSL hand shapes have meanings associated with them; for instance, of the single-extended-finger type of hand shape, all except the extended index finger have distinctive associated meanings. Thus, an extended thumb (with the exception of a few signs with body contact) indicates a male person (or sometimes gender-neutral person); the extended middle finger means *brother* (and, at least in some dialects, an allomorph for *son*, alternating with extended thumb); an extended little finger means *female person*, and the

extended ring finger is a dialectically limited alternate of the same sign with extended little finger.

The set of JSL hand shapes is subject to historic change. For instance, the *I love you* hand shape of ASL (with thumb, index, and little finger extended) originally existed only as a partial assimilation of the first morpheme to the second morpheme of the compound word *parents.* More recently, it has spread in the *I love you* borrowing, in *airplane* (and related lexemes, all originally and indigenously with the /Y/ hand shape), especially among the young, and at least in Tokyo among some young signers in the sign for *sparrow*, normally signed with the middle finger extended as well. Also, thumb extension, which has been said to be less prominent in JSL than in ASL, is more widely used in western Japan in variations of lexemes where it does not occur in eastern Japan, such as *Hokkaido* (Japan's northernmost island), *bridge*, and *sing* (all three 'normally' produced with just index and middle finger extended together).

Word Structure

It has been noted that JSL has no separate features distinguishing traditional parts of speech, such as noun from verb, or adjective from either. What *is* distinguishable are forms that can be made into predicates and forms that cannot. Perhaps better yet is the statement that JSL lexemes can be grouped into form classes, based on whether or not they can be modulated for agreement, aspect, etc.

JSL verbs are divided into those that can be inflected for person and/or number (agreement verbs), where inflection is shown by direction of sign motion and/or palm and finger orientation, and those that cannot (plain verbs). Other verbal inflections (for instance, aspect and tense) have been inadequately studied, but clearly exist. At the very least, a repeated form of a sign means a repeated, continuing action. Alternating motions, produced with one hand at a time, mean that the repetition is distributed over several participants. In general, inflection is used for aspect rather than tense, and the two signs *finish* and *middle*, which sometimes occur with verbs, are indicators of perfective and imperfective aspect rather than past and present (continuous) tense.

Nouns and pronouns are inflected for number and—unlike spoken Japanese—gender, but not case. Thus, number is indicated in nouns by reduplication, and in pronouns by multiple or sweeping pointing gestures; gender is indicated in nouns, pronouns, and verbs by the extended thumb for a *male person* and by the extended little finger for a *female person*.

Nonmanual signals, which play a major role at the syntactical level, also occur at the lexical level—for example, the tongue protruding in the cheek indicates 'falseness, lie', and often occurs as an independent gesture. The tongue-in-cheek can also occur simultaneously with another sign to indicate that what is being stated by the hands is in some way false (e.g. to sign *feigned absence from home*, one would co-sign *absent from home* along with a /tongue in cheek/).

Sentence Structure

JSL is primarily a topic-dominant language, with topic fronting being accompanied by certain nonmanual signals. In the absence of topic fronting, the semantically neutral sentence order is Subject–Object–Verb (SOV). Of the other five possible combinations of S, V, and O, only *VSO is not possible, given the proper context and topicalized element(s). Otherwise, generally the old information is given first, and the new information last.

Nonmanual signals (e.g. raising an eyebrow, tucking in the chin, nodding the head, blinking an eye, tilting the head, etc.) play a major role in syntax. For example, the difference between a wh-question—(*where is*) *Takashi*?—and a yes/no question—(*are you*) *Takashi*?—can potentially be indicated (e.g. in a one-word utterance) by a slight difference in such signals. Nonmanual markers are also used in the formation of topicalization, relative clauses, conditionals, negative sentences, and so forth. Individual nonmanual signal elements may occur in a variety of syntactic functions, alone or in combination with other nonmanual signal elements. One example of a multifunctional, nonmanual signal is tilting the head (or cocking the head), which is used for a range of modal meanings including noncommittal statements (possibility, allegedness); strangeness; approximation (in either number or location); and so forth. Perhaps all these meanings can be grouped in a unified way as deviation from the expected, stated, or normal. Nonmanual signals are an area where extensive research continues.

Lexicon and Borrowings

The overwhelming majority of JSL lexemes are indigenous. A certain percentage are shared with gestures of the majority hearing populace (e.g. *money*, *girl*(*friend*)). Borrowings from other sign languages have played a role in JSL, generally a small one, although in certain areas (e.g. computer technical vocabulary) the contribution of ASL stands out. Examples of nontechnical borrowings include *communication* and *analyze.* In earlier borrowings, ASL also served as the base for the finger-spelling syllabary—the vowels and the basic 'consonant + /a/' syllables unmarked by diacritics (e.g. *ka*, *sa*, *na*, *ha*, *ma*, *ra*, *ya*, *wa*) were borrowed from ASL vowels and consonants. The exception is *ta*, with the thumb tip provocatively protruding between closed index and

middle fingers, which is indigenous because the ASL form is taboo. Neologisms are sometimes literal translations of ASL words (e.g. *total-communication*, with /to/ -hand shape in place of ASL /t/), and the ASL method of initialism have been used (e.g. *kadai* 'subject, theme' from the /ka/ hand shape incorporated into the form of *mondai* 'problem, question'), although to a very limited extent.

Discourse Features

Discourse features such as selective use of signing space, index pointing, and role shift play a role in JSL discourse. JSL has a lexical (as well as nonmanual) topic marker *in-the-case-of*, but lacks a simple conjunctive ('and'), although it possesses a disjunctive *but* and *to-change-the-subject*, as well as conjunctions of causation *therefore* and *in-order-to*. To a certain extent, enumeration can play the role of conjunction, but this is largely left up to discourse-cohesive features such as topic continuity, sequential use of signing space, index-finger pointing (to the point in signing space assigned to a given referent or topic, or to a sign classifier already in place in that space), etc. Nonmanual signals also play a very important role in discourse topic tracking. Shared (extralinguistic) background knowledge and assumptions and linguistic indications of topic shift are very important in discourse flow, and repetition of the same sign or sign sequences are very important to ensure that signer and addressee are on the same page.

References

Fischer, Susan. 1996. The role of agreement and auxiliaries in sign languages. Lingua (Special issue on Sign Languages: Phonetics, Phonology and Morpho-syntax) 98(1–3). 103–20.

Ichida, Yasuhiro. 1991. Shuwa no Kihon-Bunpou [The basic grammar of sign language]. Shuwa-Tsuuyaku no Kisou: Shuwa-Tsuuyaku-Shi wo mezasite [The basics of sign language interpreting: aiming at being a sign language interpreter], ed. by Kanda Kazuyuki, 138–50. Tokyo: Daiichi Houki.

———. 1994. Nihon-Shwua no Bunpou to Goi [Japanese sign language grammar and vocabulary]. Nihongogaku 13(2). 25–35.

———. 1998. Nihon-Shuwa no Bunpou [Japanese sign language grammar]. Gengo 27(4). 44–51.

Minoura Nobukatsu. 1998. Nihonshuwa-hinshi-shiron — Seirisshu-shogo to no taishou wo tooshite [An essay on parts of speech in JSL: by way of comparison with the Salish languages]. Nihonshuwa-gakkai Dai 24kai-taikai Yokoushuu [Proceedings of the 24th conference of the Japan Association of Sign Linguistics]. 46–9.

Morgan, Michael. 1998. JSL ni okeru atama no naname: danwatekina kinou kara imi teki na konkyo he [Head tilt in JSL: from discourse function to semantic basis]. Nihonshuwa-gakkai Dai 24kai-taikai Yokoushuu [Proceedings of the 24th conference of the Japan Association of Sign Linguistics]. 62–5.

———. 1999. Tracking topic in Japanese sign language discourse: icon and index (summary of TISLR 7 report, in Japanese). Japanese Journal of Sign Linguistics 15(1–2). 34–5.

Osugi, Yutaka. 1999. The semantic status of 'INDEX+MALE' in Nihon Syuwa (Japanese Sign Language). Japanese Journal of Sign Linguistics 15(1–2). 1–3.

Patschke, Cynthia. 1998. A comparative study of body lean in ASL and JSL contrastive focus. Japanese Journal of Sign Linguistics 14(2). 21–39.

Smith, Wayne H. 1990. Evidence for auxiliaries in Taiwan sign language. Theoretical issues in sign language research, Vol 1: Linguistics, ed. by Susan D. Fischer and Patricia Siple, 211–28. Chicago, London: University of Chicago Press.

MICHAEL MORGAN

See also **American Sign Language**

Javanese

Javanese is one of the Austronesian languages, belonging to the Western Malayo-Polynesian subgroup and the Sundic family. The Austronesian languages exhibit a high ratio of vowels to consonants. Most root words consist of two syllables, and from these, grammatical variants are derived by means of affixes. Austronesian languages use reduplication to indicate the plural and other grammatical concepts. All these features are manifest in Javanese. Other Sundic languages are Sundanese, Tenggerese, Osing, Madurese, and Balinese, which are all spoken on or near the island of Java. An ancestor language for Javanese, Proto-Malayo-Javanic, has been reconstructed by Nothofer (1975).

Javanese is spoken by over 40% of the people in Indonesia. It is the mother tongue of 62 million people in Java, the most populous island in Indonesia. It is spoken mainly in central and eastern Java, but is also spoken in a thin strip along the north coast of west Java, except for the area around Jakarta where a form of Malay is

spoken. The regional dialect of Solo and Yogyakarta, the historical centers of Javanese culture, is called Kejawen, and is considered the standard form of Javanese.

Javanese has a literature dating back to the eighth century and its own Indian-based script, where each symbol corresponds to a syllable consisting of a consonant and a vowel. These characters may be viewed on www.skypoint.com/~gimonca/huruf-jawa.html. Old Javanese is the language of texts written in the pre-Islamic era. Nowadays, Roman script is more commonly used, but Javanese can also be written in Arabic script. The literature of Javanese has been catalogued by Pigeaud (1967).

The Javanese vowels are *a*, *e*, *i*, *o*, and *u*, and there are open(long) and closed(short) forms of each. There is also a variant of the open *a* when it is the final syllable, pronounced half way between *o* and *a*, and a neutral (pepet) *e*, as in the English word *open*. The consonants are shown in Table 1 (Robson 1992).

Root words are typically disyllables of the form (C1) V1 (C2) V2 (C3), where (C1), (C2), and (C3) are optional consonant clusters. Allowable consonant clusters include *mb*, *nd*, *ndh*, *nj*, and *nng*, which can all occur in the initial position. There is a light stress on the second last syllable, or the final syllable when the second last syllable contains a neutral *e*. This light stress does not occur when a suffix is added.

Grammatical variants of a root word may be composed by affixation, reduplication, or combination. Affixes, which may be prefixes, suffixes, or infixes, are more common in Javanese and Tagalog than in Malay. Affixes may result in the production of either a noun or a verb. Sometimes, the surface forms of affixes that result in the formation of a noun are identical to those that result in the formation of a verb. The lists of Javanese affixes used for noun and verb formation, given by Suharno (1982), are shown in Tables 2 and 3, respectively.

Adjectives can take affixes, e.g. cukup (enough) → cukupan (more or less enough), dhuwur (high) → kedhuwuren (too high). Adjectives can also be formed from nouns with affixes, e.g. jamur (fungus) → jamuren (moldy).

TABLE 1 The Javanese Consonants

	Unvoiced	Voiced	Nasal
Labial	P	B	M
Dental	T	D	N
Retroflex	TH	DH	
Palatal	C	J	NY
Velar	K	G	NG
Liquids		R	L
Semivowels		Y	W
Sibilant		S	
Aspirant		H	

There are at least six syntactic forms of word reduplication:

Whole word reduplication without any phonological change, e.g. mangan (eat) → mangan-mangan (eat informally with other people).

Partial doubling, producing a noun from an adjective, e.g. lara (sick) → lelara (sickness), peteng (dark) → pepeteng (darkness). The reduplicated fragment is a prefix consisting of the first phoneme of the root word followed by e.

Partial doubling + an, e.g. tembung (word) → tetembungan (wording, expression). Whole word reduplication of a verb with a phonological change, e.g. bali (return) → bola-bali (to and fro), mubeng (go around) → mubang-mubeng (beat around the bush).

Lexical doubling: the root words are already doubled, since the single form does not exist.

Morphological doubling—a new meaning is formed in contrast to the nondoubled one.

Robson (1992) lists the following semantic categories created by reduplication:

Do something at leisure, e.g. mlaku (walk) → mlaku-mlaku (go for a stroll).

Do something repeatedly, e.g. njerit (shriek) → jerit-jerit (go on shrieking).

Interrogative pronouns given indefinite meaning, e.g. sapa (who) → sapa-sapa or sapaa (anyone).

Mild exasperation, e.g. mentah-mentah iya dipangan (even though it is unripe he still eats it).

Plurality with diversity for both adjectives and nouns, e.g. gedhong dhuwur-dhuwur (buildings all more or less high).

Doing something together, e.g. omong-omongan (to chat together).

To compete in, e.g. gelis-gelisan (to see who is fastest at running).

Plurality, e.g. wet–wet (trees).

An important characteristic of Javanese is the speech decorum of the language, where different levels or stylemes of speech are used depending on the relative social status of the two speakers. This system has been in existence since the sixteenth century, and is a legacy of the feudal system left behind by the old Hindu court tradition. The speech levels are not different languages, but manners of speaking that vary according to the relationship between the speaker and the addressee. Each level within the language has its own characteristic set of vocabulary.

The three main levels of modern Javanese are krama, madya, and ngoko—high, middle, and low, of which krama and ngoko are most commonly used. Someone of high status speaking to someone of low status will use ngoko, while the other will use the (more formal) krama (pronounced kromo). The basic level ngoko is used between friends and equals. Ngoko means the

TABLE 2 Javanese Affixes Used in Verb Formation

Affix	Usage	Examples
moro-, mer-	Prefix forming a verb from a simple word noun. Not productive.	Dayoh (guest) → morodayoh, merdayoh (pay a visit).
kapi-	Prefix forming a verb from a simple word noun. Not productive.	Dereng (strong wish) → kapidereng (overeager).
kami-	Prefix normally occurring in conjunction with the suffix -an to form a verb from another simple word verb. Not productive.	Seset (peel) → kamisesetan (suffer from skin peel).
kumo-	Prefix forming a verb from a simple word verb or a simple word noun. Not productive.	Wani (dare) → kumowani (recklessly dare).
a-	Prefix forming a verb from a simple word noun.	Rupo (appearance) → arupo (to have the appearance of).
ma-	Prefix forming a verb from a simple word noun. This sometimes involves a sound change: a nasal consonant must be homorganic with the initial phoneme of the root word. Also applies to *n-*.	Guru (teacher) → maguru (to learn from a teacher); sembah (respect) → manembah (pay respect to); aju (progress) → maju (move forward).
n-	Prefix forming a verb from a simple word verb or simple word noun. Nasalization is found with nearly all transitive verbs. The rules for nasalization of a root word are: initial p → m; b → mb; t → n; d → nd; th → n; dh → ndh; c → ny; j → nj; k → ng; g → ngg; r → ngr; l → ngl; s → ny; w → m or ngw.	Tules (write) → nules (write); becak (pedicab) → mbecak (ride in a pedicab); gule (stew) → nggule (make stew); inep → nginep (spend the night).
ka-, ke-	Prefix forming a verb with a passive meaning from a simple-word verb.	Junjong (lift) → kajunjong (lifted); Jupoq (take) → kejupoq (taken).
di-	Prefix forming a verb with a passive meaning from a simple-word verb.	Tandor (plant) → ditandor (planted); tuku (buy) → dituku (bought).
taq-	Prefix related to *aku (I)* forming a verb with a passive meaning from a simple word verb.	Dol (sell) → taqdol (sold by me).
koq-	Prefix related to *kowe (you)* forming a verb with a passive meaning from a simple word verb.	Gawe (make) → koqgawe (made by you).
-in-	Infix forming a verb with a passive meaning from a simple word verb.	Sawang (watch) → sinawan (seen); barong (accompany) → binarong (accompanied).
-um-	Infix forming a verb from a simple word verb or a simple word noun. Usually contracted to initial *m-*.	laku (walk, gait) → mlaku (walk); kumrisik (make a rustling sound).
-r-	Infix forming a verb from a simple word noun.	Tutol (spot) → trutol (to be spotty).
-i	Suffix forming a verb from a composed-word verb, itself formed by one of the prefixes *n-*, *di-*, *taq-*, *koq-*.	Taker (measure) → nakeri (measure repeatedly); tugel (break) → ditugeli (broken into pieces); tembong (mention) → koqtembungi (you ask for a particular thing).
-an, ka- -an	Suffix or affix combination that may form a verb from a simple word verb.	Lunggoh (sit) → lungguhan (sitting); lebu (enter) → kelebon (intruded).

Continued.

TABLE 2 Continued

Affix	Usage	Examples
-en	Suffix forming an imperative verb from a simple word verb.	Jupoq (take) → jupuqen (Take it!); pangan (eat) → panganen (Eat it!).
-o	Suffix forming an imperative verb from a single word verb, or forming a hypothetical verb from a verb formed using *di-*, *tag*, *koq-*, *-in-*.	Turu (sleep) → turuo (Sleep!); gowo (carried) → gawoo (even if carried).
-ake	Suffix, always in combination with one of *ma-*, *n-*, *ka-*, *di-*, *taq-*, *koq-*, *-in-*, forming a verb from a verb.	Maguro (learn from a teacher) → maguroqake (send someone to learn from a teacher). Njupoq (take) → njupoqake (take for someone); sinugoh (served with) → sinugohake (served to somebody).

TABLE 3 Javanese Affixes Used in Noun Formation

Affix	Usage	Examples
pi-, pang-	Prefix forming a noun from a simple word verb or noun.	Anggo (use) → panganggo (clothing); utang (debt) → piutang (credit).
ka-, ke-	Prefix forming a noun from a simple word verb. In combination with the suffix *-an* it forms a verb from a simple word noun.	Weroh (see) → kawruh (knowledge); lurah (village mayor) → kelurahan (village mayor's office).
-an	Suffix forming a noun from a simple word verb or noun. It may be used in combination with *pi-* to form a noun from a verb or a noun.	Jaran (horse) → jaranan (hobby horse); tegal (nonirrigated field) → tegalan (area of such fields); Tulung (help) → pitulungan (assistance).
-e, -ne	Suffix forming a definite noun from a simple word verb or noun.	Jaran (horse) → jarane (the horse); tuku (buy) → tukune (the purchase).
-ku	Suffix related to *aku (I)*.	Kembang (flower) → kembangku (my flower).
-mu	Suffix related to *kowe (you)*.	Omah (house) → omahmu (your house).

ngoko form of *I*, while krama means *marriage*. The madya level consists of krama containing certain words shortened and with ngoko style affixes. It is often used among strangers. There are about 900 words exclusive to krama, and also a few hundred modesty words called krama inggil (inggil means *high*). These words can be mixed into either ngoko or krama as required. Krama inggil words are used when one speaks about the person, actions, or possessions of someone to whom respect is due—either addressed or referred to. These words cannot be used of oneself. Examples of the use of the three levels of Javanese speech are given by Robson.

Ngoko (girl to her younger sister): Aku wis mangan segane (I have eaten the rice).

Krama (girl to her uncle): Kula sampun nedha sekulipun (I have eaten the rice).

Krama with krama inggil (girl to her uncle about her father: Bapak sampun dhahar sekulipun (Father has eaten the rice).

Ngoko with krama inggil (girl to her sister about her father): Bapak wis dhahar segane (Father has eaten the rice).

Madya (the old servant to the girl): Kula mpun nedha sekule (I have eaten the rice).

Another form, basongan, is only used in the kratons (Sultan's palaces) of Jogjakarta and Solo. The language of religion is called 'Jawa Halus' (refined Javanese) and many words are based on Sanskrit or Kawi, but a diminishing number of people are able to use this form of the language. The number of levels may vary according to regional dialect, and between urban and rural areas (Geertz 1960). A sample of words that differ at four different levels is shown in Table 4.

The Javanese personal pronouns are shown in Table 5.

For *we*, ngoko uses *awake dhewe*, while both ngoko and krama can use the Indonesian loanword *kita*. The second- and third-person pronouns are rarely used, and are generally replaced by kinship terms, titles, or proper names. For example, a woman may be addressed as *Bu* (literally, *mother*). A pronoun may be omitted altogether if the referent's identity is understood.

TABLE 4 Words that Differ at Four Different Levels

English	Ngoko	Krama	Madya	Krama Inggil
Allow	Kareben	Kajengipun	Kajenge	Kersanipun
Obedient	Gugu	Gega	Dhaharatur	Ngestokaken dhawun
Speak	Celathu	Wicanten	Canten	Ngendika
Wear	Enggo	Engge	Ngge	Agem
You	Kowe	Sampeyan	Samangdika	Panjenengan

TABLE 5 The Javanese Personal Pronouns

English	Ngoko	Madya	Krama	Krama Inggil
I	Aku	—	Kula	Dalem
You	Kowe	Samang	Sampeyan	Panjenengan
He, She	Dheweke	—	Piyambakipun	Panjenengane, panjenenganipun

TABLE 6 Aspect Markers in Javanese

Ngoko	Krama	Meaning
Aja	Sampun	Don't
Arep, bakal	Badhe	Will
Durung	Dereng	Not yet
Isih	Taksih	Still
Lagi	Saweg	In the process of doing
Meh	—	Almost
Meksa	—	Even so, still
Ora	Mboten	Not
Padha	Sami	Also; indicates the plurality of the subject performing the action
Sok	—	On occasion, ever
Tansah	—	Always, constantly
Wis	Sampus	Already

Verbs are not inflected to denote tenses, but instead auxiliary words are used as aspect markers preceding the verb. The list given by Robson is given in Table 6.

The normal word order within the sentence is subject–predicate. There is no copulative verb, e.g. *klambiku reget* (my shirt is dirty). No changes are found in nouns or verbs for number, case, or gender.

References

Geertz, C. 1960. The religion of Java. New York: Free Press.

Pigeaud, Theodore G. 1967. Literature of Java: catalogue raisonne of Javanese manuscripts in the library of the University of Leiden, Vol. 1. Leiden: Leiden University Press.

Nothofer, B. 1975. The reconstruction of proto-Malayo-Javanic. The Hague: Martinus Nijhoff.

Robson, Stuart. 1992. Javanese grammar for students. Monash papers on South East Asia 26, Monash University, Clayton, Victoria, Australia.

Suharno, Ignatius. 1982. A descriptive study of Javanese, Pacific Linguistics, Series D, No. 45.

MICHAEL PHILIP OAKES

Jespersen, Otto

Otto Jespersen was a markedly independent scholar. He did not adopt the terminology or the viewpoints of others, he belonged to no 'school' of linguistics, and he created none, but over his long and active academic career he created an astonishing body of work in a wide variety of linguistic subfields, virtually all of which constitutes a rich and continuing source of information for later scholars and an inspiration for many.

Jespersen first became known outside his native Denmark for work in articulatory phonetics. In 1889, he developed what came to be called an antalphabetic system for describing speech sounds using letters and numbers to represent the articulating organ in the mouth (lips, tongue, etc.), the exact place of articulation in the mouth, and the degree of openness of the vocal tract in the production of a sound. The result analyzed sounds into their components in a manner more akin to modern distinctive feature analysis than to traditional phonetic transcription. For purposes of transcribing Danish and its dialects, Jespersen invented

Dania, a dialect alphabet still in use. He also published a detailed book on phonetics in Danish in 1897; it was soon translated into German, making it more widely accessible to scholars. Little was new here, but Jespersen's emphasis on first-hand observation and the scope of his examples made the texts invaluable in promoting the empirical study of speech sounds.

In a campaign to reform Danish foreign language teaching and reacting against traditional pedagogical methods that stressed rote learning, grammatical rules, and translation, Jespersen became an advocate of what he considered a more 'natural' approach, emphasizing the spoken language, the use of meaningful and interesting texts for written language, and the learning of grammar by observation and induction. *How to teach a foreign language* (1904) was influential not only in his own country but in England and particularly in North America. Indeed, Leonard Bloomfield in his most important book, *Language* (1933), presented Jespersen's 'direct' method, and elements of this later made their way into the so-called 'Army method' developed by Bloomfield and other American linguists for foreign language teaching in the years immediately preceding and then during World War II. The focus on spoken language, on meaningful material in functional context, and a more inductive approach to the acquisition of grammar all characterized both the audiolingual method used in the United States during the 1950s and 1960s and its successor, sometimes referred to as the communicative approach.

An early idea and a fundamental theme that Jespersen pursued for many years was the notion of progress in language. Based on aspects of the history of the English language, he maintained that as a language changes, original inflectional endings will disappear and be replaced by principles of fixed word order. Challenging prevailing nineteenth-century beliefs that languages decay over time, Jespersen considered such changes as progress. The modern forms, he argued, are shorter, requiring less production effort; they are fewer, requiring less memory; they are more regular morphologically and syntactically; they avoid redundancy; and their regular word order assures understanding. Jespersen admitted that his conclusions were based on information from very few languages, but nevertheless he claimed that a progressive 'tendency towards grammatical simplification is a universal fact of linguistic history' (*Language*: *its nature, development and origin,* 1921:366). Further research has supported neither the older theory of decay nor Jespersen's theory of progress.

In 1909, Jespersen began publication of what would become his most extensive study of the English language, *A modern English grammar on historical principles* (*MEG*). Contrary to an emphasis emerging at the time in both Europe and America in the work of Ferdinand de Saussure and Franz Boas, with their emphasis on synchronic study without recourse to historical information, Jespersen always maintained that the science of language was essentially historical because 'a language or a word is . . . a result of previous development and at the same time . . . the starting-point for subsequent development' (*Language*: *its nature, development and origin,* 1921:7).

The *MEG* was Jespersen's life work, his magnum opus, and he saw it as the grammatical equivalent of the *New English dictionary on historical principles*, later known as the *Oxford English dictionary* (*OED*). The volume on *Sounds and spellings* and another on *Morphology* were quite traditional and are of little interest today, but the five *Syntax* volumes remain an invaluable resource, with an abundance of citations and great insights into the nature of English syntax, cited by twentieth-century linguists as diverse as Noam Chomsky, Eugene A. Nida, and W. Nelson Francis.

In the *MEG* and in shorter books that presented similar material on syntax (e.g. *The philosophy of grammar*, 1924; *Essentials of English grammar*, 1933), Jespersen introduced the notion of ranks, a hierarchy of levels of subordination and dependency among the words of phrases (e.g. in a phrase such as *very young child*, the highest rank is held by *child*, which Jespersen would call the primary; *young* has the next highest rank, the secondary, and *very* is the lowest, the tertiary). Junction referred to subordinate constructions, nexus to predicative structures; so *the barking dog* displayed junction, *the dog is barking* nexus. Because Jespersen always held to a close connection between linguistic form and content, his grammar also made use of notional categories, universal categories of meaning.

Although work in syntax was somewhat uncommon in the linguistics of Jespersen's day, many of his ideas are precursors to aspects of modern linguistic theories. The relation of mental categories and linguistic categories is certainly central to contemporary approaches, and the ideas Jespersen sought to develop with ranks, junction, and nexus are fundamental to modern theories such as valence grammar, dependency grammar, and some types of functionalgrammar.

Jespersen's most technical work on syntax was the small book *Analytic syntax* (1937), in which he developed a notational system for representing ranks, junction, nexus, and other syntactic constructions. The system was formal and complex, and at the time, few responded to it. But as syntax and formalism became more and more a part of twentieth-century linguistics, the book was reprinted several times in the later twentieth century, a tribute to Jespersen's lifelong interest in syntax and to his lasting influence.

Unlike linguists of more modern times, Jespersen's theoretical ideas often appeared not in highly technical monographs aimed at his peers, but rather in books directed toward the educated reading public. Books such as *Language*: *its nature, development and origin* (1921), *The philosophy of grammar* (1924), and *Mankind, nature, and individual from a linguistic point of view* (1925) were described in popular journals as 'fascinating', and many members of the public came to know about linguistics by reading Otto Jespersen. The most important idea that Jespersen brought to both the public and to other linguists was the rejection of traditional grammar that set forth standards of correctness. In its place, he proposed a living grammar, based on direct observation of contemporary spoken usage, a grammar that was founded on the past but continually changing, a grammar that was systematic but not without the irregularities we should expect in anything human.

Jespersen tried to relate current usage to the users of the language. In *Language*: *its nature, development and origin*, six chapters are devoted to 'The child'. Here, Jespersen explored at length issues in the development of language in children, treating sounds, words, and grammar, as well as possible influences that children's language acquisition might have on language change. This pioneering effort to relate the nature and the universals of human language and of language change to children's language acquisition anticipated major themes in linguistics of the late twentieth and early twenty-first centuries.

The same book contained another chapter well in advance of its times, a chapter titled 'The woman', in which Jespersen presented and analyzed various accounts of differences between women's and men's speech. Modern gender studies of language use have certainly supported Jespersen's premise that women and men use language differently, often for reasons having to do with societal differences. But Jespersen's ideas were reflective of his times and place, and his chapter on 'The woman' is also a source of statements that modern scientific scholarship soundly rejects, for example, 'Men will certainly with great justice object that there is a danger of the language becoming languid and insipid if we are always to content ourselves with women's expressions . . .' (p. 247).

Jespersen's dedication to the international auxiliary language movement is sometimes treated as an odd avocation, but this interest was a rational extension of some of his most fundamental linguistic work, especially his advocacy of foreign language learning for communication, his concern with universals of language, his theories on syntax, and his principles of progress in language change. He maintained that an artificial language constructed on scientific principles should be easy to learn and efficient to use, simple in grammatical structure, the epitome of progress in language. Jespersen called his language Novial (*nov* = 'new' + *i* 'international' + *a* 'auxiliary' + *l* 'language'). But the movement was fractured and no consensus developed. Jespersen's Novial disappeared, sharing the fate of other created languages.

There is scarcely a linguistic topic of interest today that does not appear at some point in Otto Jespersen's voluminous works, which number more than 800 scholarly items in print. While some of his lines of investigation seem dated and some of his theories have been challenged by more extensive data drawn from a wider variety of languages than he used in his work, it is nevertheless often fruitful and inspiring to go back to his writings on foreign language teaching, language use, and, especially, English syntax.

Biography

Otto Jespersen was born on July 16, 1860 in Randers, Jutland, Denmark. He entered the University of Copenhagen in 1877 and studied law; in 1881, he turned to the study of languages. For seven years, while a student, he worked as a shorthand recorder in the Danish House of Parliament. He received a master's degree in 1887 with major in French, secondary concentrations in English and in Latin (the latter obligatory for language majors). He traveled for a year, meeting linguists in England (Henry Sweet, James Murray), Germany (Karl Brugmann, August Leskien, Eduard Sievers), and France (Paul Passy, Jules Gilliéron). He returned to Copenhagen, and taught English and French in private schools while working on the doctorate. He defended his doctoral dissertation 'Studies on English case' in 1891. From 1891 to 1893, he served as an unpaid Privatdocent (instructor) at the University of Copenhagen, teaching Old English and Chaucer. In 1893, he was appointed Professor of English Language and Literature, University of Copenhagen. In 1899, he was elected to membership in the Royal Danish Academy of Sciences and Letters. He visited the United States and lectured at Congress of Arts and Sciences in St. Louis in 1904, and received the French Volney Prize for *Growth and structure of the English language* in 1905. He was Dean of Faculty of Arts, University of Copenhagen from 1904 to 1906. In 1909–1910, he again visited the United States, lecturing at the University of California and Columbia University. In 1920–1921, he was Rektor (Vice Chancellor), University of Copenhagen. Jespersen retired from the University of Copenhagen in 1925, and was elected to honorary membership in the Linguistic Society of America in 1926. In 1936, he was President, Fourth International Congress of Linguists, Copenhagen. He received honorary degrees from Columbia University in

1910, St. Andrews University, Scotland in 1925, and The Sorbonne, University of Paris, in 1927. He died on April 30, 1943 in Reskilde, Denmark.

References

Aitchison, Jean. 1991. Language change: progress or decay?, 2nd edition. Cambridge, New York: Cambridge University Press.

Baker, C.L. 1995. English syntax, 2nd edition. Cambridge, MA: MIT Press.

Falk, Julia S. 1992. Otto Jespersen, Leonard Bloomfield, and American structural linguistics. Language 68.

Haislund, Niels. 1966. Otto Jespersen. Portraits of linguists: a biographical source book for the history of western linguistics, 1746–1963, Vol. 2, ed. by Thomas A. Sebeok. Bloomington and London: Indiana University Press.

Jespersen, Otto. 1889. The articulations of speech sounds represented by means of analphabetic symbols. Marburg in Hessen: Elwert.

———. 1894. Progress in language, with special reference to English. London: Sonnenschein and New York: Macmillan.

———. 1904. How to teach a foreign language. London: Sonnenschein and New York: Macmillan.

———. 1905. Growth and structure of the English language. Leipzig: Teubner and New York: Stechart.

———. 1909. A modern English grammar on historical principles, 7 vols. Heidelberg: Winter, 1914, 1927, 1931; Copenhagen: Munksgaard, 1940, 1942, 1949.

———. 1921. Language: its nature, development and origin. New York: Holt; London: Allen & Unwin, 1922.

———. 1924. The philosophy of grammar. London: Allen & Unwin and New York: Holt.

———. 1925. Mankind, nation and individual from a linguistic point of view. London: Allen & Unwin.

———. 1928. An international language. London: Allen & Unwin; New York: Norton, 1929.

———. 1933. Essentials of English grammar. London: Allen & Unwin and New York: Holt.

———. 1933. Linguistica: selected papers in English, French and German. Copenhagen: Levin & Munksgaard.

———. 1937. Analytic syntax. Copenhagen: Levin & Munksgaard and London: Allen & Unwin.

———. 1941. Efficiency in linguistic change. Copenhagen: Munksgaard.

Juul, Arne, and Hans F. Nielsen (eds.) 1989. Otto Jespersen: facets of his life and work. Amsterdam and Philadelphia: John Benjamins.

Juul, Arne, Hans F. Nielsen, and Jørgen Erik Nielsen (eds.)1995. A linguist's life: an English translation of Otto Jespersen's autobiography with notes, photos and a bibliography. Odense, Denmark: Odense University Press.

Large, Andrew. 1985. The artificial language movement. Oxford, UK and New York: Basil Blackwell.

Lightfoot, David. 1999. The development of language: acquisition, change, and evolution. Oxford, UK and Malden, MA: Blackwell Publishers.

JULIA S. FALK

Jones, Sir William

Sir William Jones, also known as 'Oriental' Jones, was a seminal linguist of the English Enlightenment, pioneering the study of comparative linguistics. Through his singular knowledge of western and oriental languages, he discovered the root relationships among them that would later be defined as the Indo-European language family. His devotion to and authoritative publications on Sanskrit and Arabic languages together with Hindu and Moslem religions and cultures made him a pioneer in introducing oriental culture to Europe. His admiration and respect for them defied and gradually began to reverse centuries of western religious and cultural ignorance and ridicule of them.

Jones studied with great success at Harrow College and Oxford University, where he held the Bennet [*sic*] fellowship. His intellectual prowess allowed him to elevate himself in society. His knowledge of the classics at Harrow so impressed Jonathan Shipley, Bishop of St. Asaph (Wales), that he recommended the young man to become the tutor in 1764 to the children of the vastly wealthy Earl of Spencer, a leading member of the independent Whig aristocrats. Tutoring financed his studies at Oxford, where he began studying the same year.

At the Spencer family seat at Althorp, he met Count Carol Revicsky, an enthusiast of Arabic and Persian, who was Polish ambassador to Great Britain and whose homeland lay in the shadow of the Ottoman Empire. Jones avidly added knowledge of these languages and cultures to the western classical ones he already knew. Arabic and Persian were of great importance in the oriental world, the first as the religious tongue of the region and the latter as its international one.

In 1770, when he was 24, Jones translated from the Persian, the *Histoire de Nader Chah*. He wrote in French, the western world's international language at the time, publishing the book in 1773 as *The history of the life of Nader Shah, King of Persia*, a tyrannical Persian ruler in India, who lived from 1688 to 1747. In 1771, Jones published *A grammar of the Persian language*, following this with pioneer translations of poetry from oriental languages.

To establish his economic and professional independence, he left the Spencer household in 1770 to study law in London at the Middle Temple. He became a lawyer and circuit magistrate in the districts of Oxford and Wales five years later and a commissioner for bankrupts. Now a noted young scholar, Jones was elected in 1773 to the Literary Club of Dr. Samuel Johnson, the renowned sage of the English language. Convening regularly at the Turk's Head Inn, club members included the historian Edward Gibbon, economist Adam Smith, painter Joshua Reynolds, and statesman Edmund Burke. A sympathizer of the movement for independence of the English colonies in America, Jones was also a friend of Benjamin Franklin, having known him from the Shipley family. Jones visited him three times in Paris after the American Declaration of Independence in 1776, where Franklin had become the young American republic's representative.

Jones's writings now advanced into the fields of law and politics. He especially concentrated on estate and commercial law, now a legal adviser to the Spencers and other aristocrats, with a work on Athenian inheritance law in 1779 and *An essay on bailments* in 1781. In addition, he wrote pamphlets and smaller works related to numerous issues regarding individual and political rights, proposing a plan for reconciliation between Britain and America based on common business interests.

To refine and deepen his oriental studies and to advance his professional position, Jones had long sought a judgeship on the bench of the English courts in India that especially oversaw operations of the East India Company. In 1783, he succeeded in being appointed an associate judge of the Supreme Court of Judicature for Bengal (now Bangladesh) in Calcutta. The same year, he was knighted by King George III, and he married Anna Maria Shipley, daughter of Bishop Shipley. They had known and been fond of each other for many years, but he vowed only to marry once he had established his financial independence. En route to the Orient, he wrote down a plan for his research, which he styled 'Objects of enquiry during my residence in Asia'. It outlined how he would survey Hindu and Moslem languages, literatures, religion, business, technology, and social and physical sciences. That he actually accomplished the study of this array of topics in the remaining 11 years of his life vaulted him to the realm of outstanding seminal scholars.

Founding the Asiatick Society in Calcutta in 1784, this association became the vehicle for propagating his voluminous research. He published *Asiatick miscellany* in 1785 and, from 1788 until his death *Asiatick researches*. He supported his research and publishing from his own financial resources. As the first president of the society, his annual presidential speeches became significant communications of his ideas and findings. The speech of 1786 (Third Anniversary Discourse) traced the common elements of western and oriental languages, the basis whereby others later identified the Indo-European family of languages. It also contributed to the development of comparative linguistics. Later discourses established the basis for establishment of comparative religious studies.

Jones also began a series of translations of epic works from oriental literature. The introduction of these classics into the English-language world and Europe would alter not only the perception of the weight and accomplishment of eastern civilizations but also influence directions and developments in modern European culture itself.

During 1784–1785 he translated nine poem-hymns to Hindu deities. In 1789, he translated *Gitagovinda* ('The song of Govinda'), a lyric poem that was one of the last Sanskrit devotional texts (*bakhti*) written. The increasing use of vernacular languages in the twelfth century began to obliterate the use of Sanskrit, somewhat as in Western Europe vernacular languages, such as Florentine Italian, began to obscure the use of Latin. In the same year, he translated the romantic verse play, the *Shakuntala* (also *Sakuntala*). This was the masterpiece of the greatest writer of classical Sanskrit, Kalidasa, a fifth-century resident of the Gupta court. In 1792, he oversaw the first printing in Sanskrit of Kalidasa's *Ritusamhara*, and translated from Arabic *Al Sirajiyyah*.

Given the vast range, quantity, and originality of the work of Sir William Jones, it seems incongruous that one can say he left his work incomplete at the time of his death in 1794. Nevertheless, he left many projects still to complete and, even more, a future promise to satisfy.

In recent decades, there has been much controversy regarding the focus and treatment of eastern cultures that Jones introduced. 'Orientalism' has been denigrated as colonialist, manipulative, and self-serving. Nevertheless, one of the central issues of modern history has been the emergence of relations between East and West as vital global concerns. It is indisputable that Sir William Jones played a key initial role in making the West aware of the East and of the need to respect the Orient's vast cultural accomplishments and to recognize the manifold characteristics that East and West share.

Biography

Sir William Jones was born in London in 1746. He was a student of Harrow College in 1753–1764, and of University College, Oxford in 1764–1768. He was private tutor to the children of the first Earl of Spencer in 1765–1770. He studied law in Middle Temple, London

in 1770–1773. Jones was elected Fellow of the Royal Society, London in 1772, and was elected as a member of the Literary Club of Samuel Johnson, London in 1773. He also practiced as a lawyer in London; as a circuit judge in Oxford and South Wales districts; and as commissioner of bankrupts in 1775–1783. Jones was knighted in 1783 and married Anna Maria Shipley the same year. He was appointed to the Supreme Court of Judicature for Bengal, Calcutta in 1783–1794. He founded the Asiatick Society of Bengal, Calcutta in 1784. Sir William Jones died as a result of liver inflammation, and was buried in Calcutta in 1794.

References

Cannon, Garland. 1990. The life and mind of Oriental Jones: Sir William Jones, the father of modern linguistics. Cambridge: Cambridge University Press.

Cannon, Garland, and Kevin R. Brine (eds.) 1995. Objects of inquiry: the life, contributions, and influences of Sir William Jones (1746–1794) [Papers from Sir William Jones Symposium on the occasion of bicentenary of birth, New York University, April 21, 1994]. New York: New York University Press.

Franklin, Michael J. 1995. Sir William Jones. Writers of Wales series. Cardiff: University of Wales Press.

———. 1995. Sir William Jones: selected poetical and prose works. Cardiff: University of Wales Press.

Jones, Sir William. 1785. Asiatick miscellany.

———. 1788–1794. Asiatick researches.

———. 1771. Dissertation sur la littérature orientale.

———. 1781. Essay on the law of bailments.

———. 1789. Gita Govinda by Jayadeva, translation from Sanskrit.

———. 1771. Grammar of the Persian language.

———. 1770. Histoire de Nader Chah, translated from Persian; translated into English as The history of the life of Nader Shah, King of Persia. 1773.

———. 1789. Shakuntala by Kalidasa, translation from Sanskrit.

———. 1807. The collected works of Sir William Jones, 13 vols. London: J. Stockdale; reprinted, London: Curzon; and New York: New York University Press, 1993.

Said, Edward. 1979. Orientalism. New York: Random House.

EDWARD A. RIEDINGER

Juba and Nubi Arabic

Juba and Nubi Arabic are two contemporary related Arabic-based Pidgins–Creoles descending from an Arabic military pidgin that developed in southern Sudan following the annexation of Sudan by the Turkish–Egyptian Government in 1820. While the northern and central parts of Sudan were already largely Arabized at the end of the eighteenth century, the southern part of Sudan laid behind the southern borders of Arabic expansion due to physical barriers. Thus, the conquest of southern Sudan between 1839 and 1841 instituted the first trade contacts between the north and the southern hinterlands. The subsequent establishment of military and trade camps and the development of a large-scale slave trade between 1854 and 1889 led to major social upheavals and to the emergence of an Arabic pidgin in this previously non-Arabic-speaking area. Many members of the local southern population (sometimes up to 25%) were captured as slaves or employed in the camps, whose population could number up to 10,000 people. In these camps, the southern population soon outnumbered the Arabic-speaking population. Coming from heterogeneous linguistic and ethnic backgrounds, this southern detribalized population quickly adopted a pidginized form of Arabic as its lingua franca. The exact origin and form of this Arabic pidgin remains unknown. Its target language seems to have been a mixture and an approximate form of Egyptian and northern/western Sudanese Colloquial Arabic. It is not certain whether a single variety spread all around southern Sudan or whether various varieties emerged simultaneously in the different camps. But the available historical and contemporary data testify to a very similar restructuring process and indicate that this Arabic-based pidgin stabilized in a very short span of time.

Following various military events at the end of the nineteenth century, the southern Sudanese Arabic pidgin first known as Bimbashi Arabic (from Osmanli, Bimbashi 'officer') developed in three further varieties: Turku in Chad, Nubi in Kenya and Uganda, and Juba Arabic in southern Sudan. Turku (from Arabic, Turuk, i.e. 'Turks') was brought from Bahr al Ghazal to Chad at the end of the nineteenth century by some Sudanese soldiers from Rabeh's army. It remained a pidgin and became one of the trade languages of Chad and neighboring countries in the first decades of the twentieth century. No present-day Turku is recorded; also, some kind of Arabic functions as interethnic lingua franca in this area. Nubi (or Ki-Nubi) was brought to Kenya–Uganda in 1888 by the retreating Sudanese

567

soldiers of Amin Pacha's army. Those Arabized and Islamized soldiers stayed and formed a new ethnic groups till locally known as Nubi and speaking an Arabic-based Creole, ki-Nubi, as their mother tongue. Cut off from further Arabic influence, Ki-Nubi developed in a dominant African/Swahili environment and is not mutually intelligible with standard Arabic (both Colloquial or Classical Arabic). Most East African Nubis are urban dwellers and are at least bilingual. In southern Sudan and more specifically in Equatoria, Bimbashi Arabic continued to develop during the first decades of the twentieth century first in the military garrisons and urban centers and then spread as the main lingua franca in the heterogeneous rural areas with minimal contact with standard Arabic. Today, it is becoming the mother tongue of a growing number of children in southern urban areas due to mixed neighboring and interethnic marriages. It thus functions both as a first and second language. It became known as Juba Arabic from Juba, the capital of Equatoria. Since 1956 and more intensively since the 1980s, Juba Arabic speakers have been more exposed to Colloquial Sudanese Arabic and Classical Arabic through mass media, schooling, urbanization, and migration to northern Sudan. Therefore, many Juba Arabic speakers can shift from a more Creolized variety to a more Colloquial one according to the setting. However, the antagonistic relationship between the North and the South fosters the identity function of Juba Arabic, which is perceived as expressing and symbolizing an 'African identity'.

Ki-Nubi and Juba Arabic are mainly spoken languages. They are not taught and have neither been standardized nor normalized through an official script. Ki-Nubi is an ethnic language for approximately 15,000 Kenyan speakers and an unknown number of Ugandans. Juba Arabic, on the other hand, is both a vehicle for the majority of the southerners and a vernacular for some urban dwellers. Its formal contexts of use include limited radio broadcastings, theatrical performances, songs, Christian religious preaching, etc. Some prayer books are written in Juba Arabic using Latin script, while individuals educated in Literary Arabic may use Arabic script. Both Nubi and Juba Arabic include regional varieties. But Nubi appeared more homogenized than Juba Arabic, which encom-

passes a wide range of individual variations. In spite of a short time of common history and a century of separate development, Nubi and Juba Arabic are structurally closed and mutually intelligible, which may indicate that their common ancestor stabilized before their split. The lexicon derives mainly from Arabic roots, but with important phonological restructuring and the addition of a number of borrowed words from African vernaculars. The morphology of Arabic has been lost (such as verbal flexions, derived verbal and nominal forms). Both languages used invariable verbal and nominal stems and independent grammatical markers for expressing grammatical categories such as tense, aspect, persons, definiteness, comparative, etc. Like for most Pidgins and Creole languages, the attention of linguists focused on the supposed origin of these linguistic features. In this respect, Arabic-based Pidgins–Creoles appear to share many similarities with other non-Arabic-based Pidgins–Creoles especially regarding the verbal system. Their restructuring has gone further than any other Arabic Colloquial varieties, including the most peripheral ones, and interferences from neighboring African languages have been pointed out at various levels. The three above-mentioned Arabic-based Pidgins–Creoles might not have been the only Arabic-based existing ones but their history testified to a radical context of emergence compared to other situations of contact. The maintenance of Nubi and Juba Arabic as specific Pidgin/Creole varieties also indicates the crucial role played by identity factors.

References

Heine, Bernd. 1982. The Nubi language of Kibera: an Arabic Creole. Berlin: Dietrich Reimer.

Kaye, Alan, and Mauro Tosco. 1993. Early East African Pidgin Arabic. SUGIA 14. 269–305.

Mahmud, Ushari. 1982. Arabic in the Southern Sudan: history and spread of a Pigin–Creole. Khartoum: FAL.

Owens, Jonathan. 1985. The origin of East African Nubi. Anthropological Linguistics 27(3). 229–71.

———. 1997. Arabic-based Pidgins and Creole. Contact languages, a wider perspective, ed. by Sarah G. Thomason. Amsterdam, Philadelphia: John Benjamins Publishing. 125–72.

Tosco, Mauro, and Jonathan Owens. 1993. Turku: a descriptive and comparative study. SUGIA 14. 177–268.

CATHERINE MILLER

K

Kayardild and the Tangkic Languages

The Tangkic languages are spoken in Queensland, Australia, in the Wellesley Islands and adjoining mainland. The family comprises three main languages: Lardil, Kayardild, and Yukulta. Additional and now extinct varieties are Yangkaal and Nguburindi, although the limited materials we have on these show Yangkaal to be a sister dialect of Kayardild and Nguburindi a sister dialect of Yukulta. Within Tangkic, the family bisects into 'northern Tangkic', comprising Lardil alone, and 'southern Tangkic', comprising Yukulta and Kayardild.

The Tangkic languages have no close relatives, although they are related, at a distant level, to other Australian languages and share most grammatical similarities with languages along the Roper River, well to the west.

Speakers of the Tangkic languages were traditionally hunter–gatherers, with a strong emphasis on marine resources, building stone walls around the coasts to catch fish and hunting for the abundant turtle and dugong found in the area. Apart from the Kayardild, who were relatively isolated, the other Tangkic groups were linked together in a complex network of tribal interrelationships with people to the west and south. From the 1860s for the Yukulta, the 1920s for the Lardil, and the 1940s for the Kayardild, they came into increasingly intensive contact with pastoralists on the mainland and missionaries on the islands, which disrupted many aspects of traditional life, although there has been a recent resurgence in many traditional activities. However, no Tangkic language is now spoken fluently by people below the age of 40 and none has more than a dozen speakers left: English has replaced them, albeit in a distinctive form containing many words adapted from Tangkic languages.

Apart from scanty word lists from the nineteenth and early twentieth centuries, all materials on the Tangkic languages have been recorded since the early 1960s. The practical orthographies (written language systems) used to write these languages were developed in this period. These orthographies use digraphs for a variety of sounds; e.g. *r* before a stop or nasal letter denotes retroflexion. For example, *rd* or *rn* indicates that the *d* and *n* are pronounced with the tip of the tongue curled backward. Similarly, an *h* after a stop (*th*) or nasal letter (*nh*) indicates that the blade of the tongue is placed between the teeth for the pronunciation of the relevant sounds. As far as their sound system is concerned, the Tangkic languages are typical Australian languages.

Lardil differs from the other Tangkic languages in having lost the final syllable in words greater than two syllables: cf. Kayardild *kandukandu,* 'red', but Lardil *kandukan.* Furthermore, Lardil dropped off any remaining word-final *ng*, *k, m,* or *b*: cf. Kayardild *kurkangka,* 'bulrush', but Lardil *kurka* (rather than *kurkang*). Because various grammatical suffixes 'protect' the original word stem from such truncation, Lardil has developed a systematic distinction of word forms, depending on their function and position in the sentence. The Tangkic languages are typical of Australian languages in using a rich system of case suffixes, i.e. word endings that make distinctions similar to English *the house* vs. *to the house.* This allows for great freedom of word order. Beyond this, their case systems are perhaps the most remarkable in the

world for several interrelated reasons. First—and by no means uniquely in Australia—they exhibit *double case marking,* because one noun phrase embedded in another inflects both for its own case (e.g. the possessive) and that of the main noun: cf. Kayardild *thabuju-karra,* 'brother's', and *wangal-nguni,* 'with the boomerang' (literally 'boomerang-with'), combine to *thabuju-karra-nguni wangal-nguni,* 'with brother's boomerang' (literally 'brother's-with boomerang-with'). Second, Kayardild and Lardil mark e.g. tense, not only on the verb but simultaneously via *modal case* suffixes on nouns. Thus, Ksayardild *dangka-a burldi-jarra yarbuth-ina thabuju-karra-nguni-na wangal-nguni-na,* 'the man hit the bird with brother's boomerang', shows a past-tense suffix *-jarra* on the verb *burldi,* 'hit', and also a modal case suffix *-(i)na* on the nouns *yarbuth,* 'bird', *thabuju-karra-nguni,* 'with brother's', and *wangal-nguni*, 'with the boomerang'. Third, in Kayardild, case suffixes can be used to connect sentences with each other. In this usage, the relevant suffix occurs on all words of the subordinate (dependent) clause. Much of how this strange system evolved has now been reconstructed with the help of data from Yukulta, the Tangkic language that conserves the most characteristics of the common ancestor language.

Another twist to case in the Tangkic languages is added by a further set of *verbal case* inflections, semantically and structurally part of the set of normal case inflections, but with the peculiarity that they convert their hosts from nouns into verbs. Kayardild *ngada waa-jarra wangarr-ina ngijin-maru-tharra thabuju-maru-tharra,* 'I sang a song for my brother', for example, contains *thabuju-maru-tharra,* 'for my brother', which literally means 'brother-put-(past tense)'. Thus, the *verbal dative case* suffix *-maru-* converts the noun *thabuju* into a verb, which then takes the regular verbal inflection for past tense: *-tharra.*

Like many other Australian languages, the Tangkic languages have a highly developed set of derivatives from compass terms. To locate an entity, one normally says things such as 'the east uncle' or 'the shark coming from the east'; some Kayardild examples of derivatives based on the root *ri-*, 'east', are *riinda,* 'coming from the east', *rilungka*, 'eastward', *riliida*, 'heading ever eastwards', *riyananganda*, 'to the east of', *ringurrnga*, 'east across a geographical boundary', *riyanyinda*, 'at the eastern extremity of', *rilumir-damirda,* 'sea-grass territory to the east', *rilurayaan-da,* 'from one's previous night's camp in the east', *rilijulutha,* 'move to the east', and *riinmali,* 'hey you coming from the east'.

Lardil and Yangkaal people had a special auxiliary language, known as Damin, taught to second-degree initiates, which involved substituting all roots (but not grammatical suffixes) with special forms drawn from a set of approximately 150 items with abstract meaning and a bizarre sound structure. In fact, the sound structure of Damin is unique among the world's languages in using five distinct airstream types: some sounds are produced with the usual mechanism of pressing air outward from the lungs (pulmonic egressive), others are produced with inward breath (pulmonic ingressive), and others are click sounds (velaric ingressive) and sounds produced by ejecting air via tongue movements (labiovelar lingual egressive) and by movements of the Adam's apple (glottalic egressive). Of these, the pulmonic ingressive and labiovelar lingual egressives are found nowhere else in the world's languages, at least not with a similarly prominent function. As far as meaning is concerned, Damin solves the problem of compressing the entire everyday language vocabulary into 150 terms by four methods:

- Highly abstract words, such as a single form for 'act harmfully/damagingly upon' to replace the everyday words for 'eat', 'chop', 'bite', 'shoot', and 'cut'.
- A variety of semantic extensions are used to tie together many distinct senses for the relevant word; e.g. 'stomach' is also used to mean 'large intestine' (by contiguity), 'guts, excrement' (by association), 'defecate' (product to process), and 'emerge' (by metaphor).
- The same word can be used as a noun or a verb, as shown by the extension from 'excrement' to 'defecate', above; inflectional suffixes show which is intended.
- The Damin words are supplemented by hand signs to distinguish subtypes, e.g. the abstract term for 'fish' means 'bluefish', which crunches coral with its teeth, if the speaker points to the teeth. Because these speech–sign combinations are conventionalized and taught together, Damin should probably be viewed as a mixed speech and sign language.

References

Dench, Alan, and Nicholas D. Evans. 1988. Multiple case-marking in Australian languages. Australian Journal of Linguistics 8. 1–47.

Evans, Nicholas D. 1992a. Kayardild dictionary and ethnothesaurus. Melbourne: University of Melbourne Department of Linguistics.

Evans, Nicholas D. 1992b. 'Multiple semiotic systems, hyperpolysemy and the reconstruction of semantic change in Australian languages. Diachrony with synchrony: language history and cognition, ed. by Günter Kellermann and Michael D. Morrissey, 475–508. Frankfurt: Peter Lang.

Evans, Nicholas D. 1995. A grammar of Kayardild, with historical-comparative notes on Tangkic. Berlin: Mouton de Gruyter.

Hale, Kenneth L. 1973. Deep-surface canonical disparities in relation to analysis and change: an Australian example. Current Trends in Linguistics 8: Linguistics in Oceania, ed. by Thomas A. Sebeok, 401–58. The Hague: Mouton.

Hale, Kenneth L. 1982. The logic of Damin kinship terminology. Languages of kinship in Aboriginal Australia, ed. by Jeffrey Heath, Alan Rumsey, and Francesca Merlan, 31–37. Sydney: Oceania Linguistic Monographs.

Hale, Kenneth, and David Nash. 1997. Damin and Lardil phonotactics. Boundary Rider: essays in honours of Geoffrey O'Grady, ed. by Darrell Tryon and Michael Walsh, 247–59. Canberra: Pacific Linguistics.

Keen, Sandra. 1983. Yukulta. Handbook of Australian languages, Vol. 3, ed. by R.M.W. Dixon and Barry J. Blake, 190–304. Canberra: Australian National University Press.

McConvell, Patrick. 1981. How Lardil became accusative. Lingua 55. 141–79.

McKnight, David. 1999. People, countries and the Rainbow Serpent. New York, Oxford: Oxford University Press.

Ngakulmungan, Kangka Leman. 1997. Lardil dictionary. Gununa, Queensland: Mornington Shire Council.

Plank, Frans (ed.) 1995. Double case: agreement by Suffixaufnahme. Oxford: Oxford University Press.

NICHOLAS EVANS

Khmer and Mon-Khmer Languages

The largest of the branches making up the Austroasiatic language family, the Mon-Khmer language group comprises over a hundred languages spoken throughout Southeast Asia, particularly in the countries of Kampuchea, Vietnam, and Malaysia. While primarily concentrated on the mainland of Southeast Asia, Mon-Khmer languages can be found as far west as the Republic of India and on various islands surrounding the Asian mainland. The distribution of Mon-Khmer languages throughout Southeast Asia suggests that they were among the first to be spoken in the region. Although it is hard to be certain without written evidence, some scholars have speculated that the branches of Mon-Khmer separated at approximately the same time during the second millennium BC. The most commonly accepted subgrouping of the Mon-Khmer languages proposes a classification of nine branches: Khmer, Monic, Pearic, Bahnaric, Katuic, Khmuic, Palaungic, Khasi, and Viet-Mu'o'ng.

Mon-Khmer languages possess several characteristics distinguishing them from other languages spoken in Southeast Asia. Among the most significant of these is that words in Mon-Khmer languages tend to be sesquisyllabic, i.e. a word consists of an initial unstressed syllable followed by a stressed syllable with full vocalism, in contrast to the high frequency of monosyllabic and disyllabic words in other languages in the region. Another characteristic making Mon-Khmer languages unique to the region is that they generally do not recognize tone as phonemic, i.e. creating a difference in meaning, in contrast to tonal languages in the region such as Thai, Lao, and Chinese. Mon-Khmer languages do, however, make a phonemic distinction between such vowel qualities as breathiness and creakiness, and the aspiration of consonants is recognized as phonemic as well. An additional characteristic that makes Mon-Khmer languages unique is that they possess very large, complex vowel systems. For example, the most complex vowel system among these languages is that of Bru, a member of the Katuic branch of Mon-Khmer, which has 68 vowels in its inventory. Conversely, the inventory of consonants in Mon-Khmer languages is relatively small, due in part to the absence of voiced sounds found in many other languages of the world.

Khmer

One of the oldest languages in Southeast Asia, Khmer is the official language of Kampuchea (formerly Cambodia). In addition to the six million Kampucheans who speak Khmer, political turbulence in Kampuchea, especially during the 1970s, resulted in many people fleeing the country and creating sizable populations of Khmer speakers outside Kampuchea, primarily in the neighboring countries of Thailand, Vietnam, and Laos. Khmer speakers also reside in France and the United States, bringing the total number of Khmer speakers to over seven million. Modern Standard Khmer is used throughout Kampuchea; however, regional differences are typically slight enough that there are few obstacles in cross-dialectal communication.

As a result of thousands of years of contact between various populations in Southeast Asia, many of the languages spoken in the region share much of the same vocabulary. During the Angkor period from the ninth to the fifteenth centuries CE, Khmer had a powerful influence on other languages in the region,

an influence that extended to unrelated languages such as Lao and Thai, as well as to other Mon-Khmer languages, such as those of the Pearic branch, which includes Chong and Pear, and the Bahnaric branch, which includes Phnong and Chrau. Khmer and Thai, which is primarily spoken in neighboring Thailand, share so much with respect to vocabulary that early scholars were led to believe that the two languages were genetically related. It is now assumed, however, that the similarity between the two languages is due to mutual borrowing and reborrowing between them. This borrowing has extended beyond lexical borrowing to include syntactic borrowing as well. The Khmer lexicon also includes borrowings from Vietnamese.

Khmer has also borrowed extensively from languages outside the region of Southeast Asia. During the Angkor period, Khmer borrowed much of its institutional vocabulary from Sanskrit and some of these borrowings can now be heard in colloquial speech, while others are typically only found in literary works. Since the introduction of the Buddhist religion to Kampuchea in the fifteenth century, Khmer has also borrowed heavily from Pali. During the political domination of Kampuchea by the French from the mid-nineteenth to the mid-twentieth century, Khmer borrowed many French words, some of which can still be heard, particularly in Kampuchea's urban centers. Khmer has also borrowed a small number of Chinese words. Since Khmer words are generally monosyllabic and sesquisyllabic, polysyllabicity is typically indicative of borrowings, most often from Pali, Sanskrit, or French.

Like other Mon-Khmer languages, the inventory of sounds in Khmer is interesting, mostly owing to its large vowel system. While not as vocalically rich as Bru, most standard dialects of Khmer have between 25 and 28 vowels, with a typical inventory comprising 27 vowels. The Northern Khmer dialect has a system comprising at least 31 vowels. The consonantal inventory of Khmer is somewhat smaller than its vocalic inventory, partially due to the absence of voiced sounds such as *g* and *z* that speakers of other languages would expect to find. Phonologically, consonant clusters may occur at the beginning of a Khmer word, but not at the end. Additionally, there exist several consonants never found word-finally in Khmer, including *r, s, b,* and *d.*

Like Chinese and Vietnamese, Khmer is an isolating language, which means that it tends toward a one-to-one correspondence between morpheme and word. Khmer never uses inflectional morphology, but relies on separate morphemes or context to show such things as gender, number, and tense. Although lacking inflec-

tional morphology, Khmer uses derivational morphology to create new words. With respect to affixing, Khmer does not use suffixing, but uses prefixing and infixing, particularly as a means of deriving nouns from verbs. Khmer also uses compounding extensively and reduplication is common, particularly with vowel alternation in the prefixed reduplicant to create a class of words known as iconic expressives.

The basic word order of Khmer is subject–verb–object, although topicalization, or the movement of an object to the beginning of a sentence, is a common phenomenon in Khmer. Questions retain the same word order as statements in Khmer. Adjectives and other elements dependent on the noun generally follow the nouns they modify, an exception being that numerals typically precede the noun. Negation is preverbal with an intensifier typically following the verb.

Mon

Prior to the sixteenth century CE, the Mon ruled over parts of Myanmar and Thailand, but today the Mon live primarily in small villages in Myanmar (formerly Burma) and in Thailand. In Myanmar, they generally live along the Tenasserim coast between Thaton and Tavoy, and also live in the vicinity of Pegu. Although the Mon have made many efforts to attain recognition as a state in Myanmar, these attempts have met with little success. As a result, over the past several centuries many Mon have fled to Thailand, where they live primarily in the Chao Phraya basin between Uthiathani and Bangkok. Although they do not live in the city of Bangkok itself, the Mon have established a cultural center in Thailand on the southern outskirts of the city in the town of Prapradaeng. Despite this success, the southern portion of Myanmar is still considered the homeland of the Mon.

The language of the Mon has a story parallel to the story of its speakers. A member of the Monic branch of Mon-Khmer languages, Mon was once a major language in Southeast Asia; however, it is now spoken by fewer than a million people. Although for centuries there has been pressure on the Mon in both Myanmar and Thailand to assimilate to the dominant languages and culture in the countries—Burmese and Thai, respectively—the pressure to assimilate has increased since World War II due to modernization and urbanization. In Thailand, where the Mon have been living in exile for centuries, there is intense pressure for all ethnic minorities to assimilate—the Thai government makes no distinction between ethnic groups and Thai. Additionally, the Thai language and culture are the only local language and culture recognized by the government schools. Not surprisingly, the Mon exiles

have made no attempts to gain autonomy in their adopted country of Thailand.

Khasi

The westernmost language in the Mon-Khmer group, Khasi is spoken by approximately 500,000 people living in the Khasi and Jaintia Hills of Meghalaya; as such, it is the only Mon-Khmer language found in the Republic of India. Khasi is also spoken by about 85,000 people in Bangladesh. Because of its geographical isolation from other Mon-Khmer languages, Khasi is the sole member of the Khasi branch of the Mon-Khmer group. Its status as a singular language in India, as well as its strong matrilineal society, has attracted a great deal of interest in the Khasi people, particularly among ethnographers and other scholars. As a result, the Khasi people have managed to maintain some prestige in the region.

Vietnamese

Vietnamese is the official language of Vietnam and comprises approximately 80 million speakers, primarily in Vietnam, but also in neighboring countries, as well as in France and the United States. Although Vietnamese is now generally recognized as being a member of the Mon-Khmer family, its genetic affiliation was the subject of debate among linguists for some time. The controversy arose due to the status of Vietnamese as a tonal language, a trait also possessed by Chinese and Thai, but a trait not shared by the majority of Mon-Khmer languages. Although tone is not the only characteristic that Chinese and Vietnamese share, another being the great number of lexical items the two hold in common, linguists ruled out any close genetic affiliation between the two languages long ago, instead attributing similarities, particularly with respect to vocabulary, to the long domination of the Vietnamese by the Chinese. Given the choice between Vietnamese as a Mon-Khmer language that had somehow adopted tone, or a Tai language that had borrowed much of its vocabulary from Mon-Khmer languages, many linguists adopted the latter view until the convincing argument of Haudricourt (1954). A French botanist, Haudricourt argued that the phenomenon of tonal genesis in Vietnamese was an evolutionary process in which the language acquired three tones by the sixth century through language contact with speakers and through internal changes, and doubled this number by the twelfth century to the six tones currently used in the Hanoi dialect of Vietnamese. Haudricourt went on to argue that the tones in Chinese and Thai developed in much the same way.

Despite their differences, Vietnamese and other Mon-Khmer languages share a number of characteristics. Vietnamese is an isolating language with no inflectional morphology; however, like other Mon-Khmer languages, Vietnamese uses derivational morphology as a means of creating new words. Besides prefixing, Vietnamese makes extensive use of compounding and also uses reduplication, particularly in onomatopoetic expressions.

Among the Austroasiatic languages, three have long traditions of writing: Khmer, Mon, and Vietnamese. The earliest Khmer writings in which the date is known were written in 611 CE, and other extant writings are estimated to be from the second or third centuries CE. Khmer is written in a phonologically based writing system that derives from the Pallava script, a system that was used in South India at approximately the same time and descended from the Bhrami script. Having evolved since its introduction, the script used for writing in Khmer is now called the *Khmer letters*. Mon is also written with an Indian-derived alphabet and extant texts include some dating back to the eleventh century CE. In addition, epigraphs have been found that date back to approximately 600 CE. Although Vietnamese also has a long written tradition, the history of this writing is considerably different from the histories of writing in Khmer and Mon. Written Vietnamese begins with Chinese characters, then with a writing system derived from Chinese characters, and finally a Roman script introduced by Catholic missionaries in the seventeenth century. Although scholars have speculated that the Vietnamese were using the Chinese script to record their own language as early as the eighth century CE, the earliest extant writings of Vietnamese in this script date back to the thirteenth century.

References

Haudricourt, André-Georges. 1954. De l'origine des tons in viêtnamien. Journal Asiatique 242. 68–82.

Huffman, Franklin E. 1970. Cambodian system of writing and beginning reader with drills and glossary. New Haven: Yale University Press.

Jacob, Judith M. 1968. Introduction to Cambodian. Oxford: Oxford University Press.

Parkin, Robert. 1991. A guide to Austroasiatic speakers and their languages. Honolulu: University of Hawaii Press.

Thomas, David D. 1964. A survey of Austroasiatic and Mon-Khmer comparative studies. Mon-Khmer studies, Vol. I. Saigon: The Linguistic Circle of Saigon.

Thomas, David D., and Robert K. Headley, Jr. 1970. More on Mon-Khmer subgroupings. Lingua 25. 398–418.

LAMONT ANTIEAU

See also **Austroasiatic; Southeast Asia; Vietnamese**

Khoisan

According to a widely accepted genetic classification by Joseph Greenberg (1963), Khoisan constitutes one of the four language phyla (or families) to be found on the African continent. Apart from two languages (Sandawe, Hadza) spoken in north-central Tanzania, all languages are found in southern Africa, mostly in Botswana and Namibia, and to a minor extent also in Angola, Zambia, and South Africa. While there must have been more than 100 Khoisan languages prior to the European conquest of southern Africa, less than 30 have survived. The total number of Khoisan speakers probably does not exceed 200,000, of whom around two thirds belong to one language community: Nama/Damara (cf. Güldemann and Vossen 2000).

Earlier work on Khoisan had assumed, largely based on nonlinguistic criteria taken from physical and cultural anthropology, that there exists a basic division between 'Hottentot' and 'Bushman' languages. The former were said to consist of the Khoekhoe branch of the Khoe family (see Table 1). The latter includes the remainder of Khoisan, with the exception of the Tanzanian languages Sandawe and Hadza, which were not associated with this general entity at that time (see e.g. Meinhof 1912). This division is reflected in the label 'Khoisan', proposed by Schultze (1928), which is composed of two lexical items from the Nama/Damara language: *khoi* (or *khoe*) 'person', standing for the 'Hottentot', and *san* 'forager', standing for the 'Bushman' languages.

Subsequent research established that there exists some kind of linguistic relationship uniting the Khoisan languages of southern Africa (cf. Köhler 1975). Greenberg (1963) was the first to argue that Khoisan forms a genetic unit, and that this unit not only includes the 'click languages' of southern Africa but also two East African languages. He proposed the following classification of Khoisan:

(1) South African Khoisan
 a. Northern
 b. Central
 c. Southern
(2) Sandawe
(3) Hatsa (= Hadza)

A number of Khoisanists are reluctant to adopt Greenberg's (1963) hypothesis that the Khoisan languages really are genetically related to one another. Greenberg's phylum hypothesis has been defended by some (e.g. Honken 1977; Ehret 1986) but rejected by others (e.g. Westphal 1971). A third group of scholars (e.g. Köhler 1981; Sands 1998; Traill 1986; Güldemann and Vossen 2000) do not explicitly reject Greenberg's hypothesis but argue that so far there has been no proof of genetic relationship; given the evidence available, this appears to be the most reasonable position to be assumed.

A substantial number of lexical and grammatical similarities across various groupings of Khoisan have been identified, but in most cases it remains unclear whether these similarities are due to genetic relationship or language contact. Only for one of the Khoisan families, Khoe, there exists a sound comparative analysis (Vossen 1997).

Table 1 gives an overview of Khoisan languages and their genetic classification (for more details and a slightly different treatment, see Güldemann and Vossen 2000). Some of the languages are presumably extinct; others form dialect clusters. The !Xun family, for example, constitutes a complex dialect cluster connected by chains of mutual intelligibility, even if many East !Xun speakers do not understand North !Xun speakers. In most cases, it remains unclear whether there is a language or a dialect boundary. For example, Kxoe, Buga, and ǁAni are treated here as different languages of Kalahari Khoe; yet, there are also reasons to consider them as dialects of one and the same language (see Table 1).

While it remains unclear whether indeed the Khoisan languages form one language family, there are a few typological characteristics that set these languages off from the neighboring Bantu languages in particular and from many other languages in general:

(1) Most Khoisan languages distinguish between at least four different types of click; languages of the !Ui-Taa and the ǂHõã families even have five. In addition to dental (ǀ), alveolar (!), palatal (ǂ), and lateral (ǁ) clicks, there is a bilabial click (ʘ) in these two families.
(2) Furthermore, there are a number of distinctions made in vowels; !Xun, for example, has distinct sets of oral, nasal, glottalized, and pharyngealized vowels.
(3) Accordingly, the number of phonemes in these languages is comparatively high; some of these languages, such as !Xõõ and !Xun, have more than one hundred distinct phonemes and thus

TABLE 1   Khoisan Languages
DC = dialect cluser, C = central, E = east, N = north, S = south, W = west, ⁺ = presumably extinct language.

Family	Branch	Group	Languages	Where Spoken	Number of Speakers
!Xun (Northern)			North !Xun (DC)	S Angola, N Namibia	Less than 5,000
			East !Xun (Juǀ'hoan; DC)	NE Namibia, NW Botswana	Less than 5,000
Khoe (Central)	Khoekhoe	North	Namara/Damara, Haiǀǀ'om, ⧺Aakhoe (DC)	Namibia	Over 120,000
		South	⁺!Ora; ⁺Cape varieties (DC)	S Namibia, South Africa	Presumably extinct
	Kalahari Khoe	West	Kxoe, Buga, ǀǀAni (DC); Naro (DC); Gǀǀana, Gǀui, ⧺Haba (DC)	NE Namibia, Botswana, SE Angola	Over 10,000
		East	Shua, Ts'ixa, Danisi, ǀXaise, ⁺Deti; Kua-Tsua (DC)	Botswana	Less than 5,000
!Ui-Taa (Southern)	!Ui		⁺ǀXam, ⁺ǀ'Auni, ⁺⧺Khomani, ⁺ǀǀXegwi, etc.	South Africa, E Namibia, W Botswana	Presumably extinct
	Taa		!Xõõ (DC)	W Botswana	A few thousand
⧺Hõã			⧺Hõã	WC Botswana	A few hundred
Kwadi			⁺Kwadi	SW Angola	Presumably extinct
Sandawe			Sandawe	C Tanzania	A few ten thousands
Hadza			Hadza	C Tanzania	A few hundred

belong to the phonetically most complex languages of the world.

(4) Although Khoisan languages have a fairly rich inventory of grammatical distinctions, their morphological structure is fairly isolating analytic.

A widespread morphological characteristic of Khoisan languages is the presence of a noun class system. There is, however, a remarkable difference between the Khoe languages on the one hand, and the !Xun and Taa languages on the other. While the latter do not make a morphological distinction based on natural sex, marking distinctions such as human vs. non-human or animate vs. inanimate instead, Khoe languages have three gender categories, which are masculine, feminine, and common. Khoe languages are further characterized by a set of portmanteau morphemes simultaneously expressing person, gender, and number (pgn-markers); since there are three categories of each person, gender, and number (singular, dual, plural), there are close to 27 different pgn-markers, as well as personal pronouns, in these languages. Sex-based gender systems are also found in Kwadi, Sandawe, and Hadza.

Widespread in the verbal system (except for Khoe) are suppletive verbs, where there are different forms for singular and plural. Suppletive verbs show number agreement with the object in the case of transitive verbs and with the subject in the case of intransitive verbs. Furthermore, a number of Khoisan languages have verbal derivational morphemes, invariably suffixes or enclitics on the verb.

Most Khoisan languages have subject–verb–object (SVO) as their basic word order, the noun precedes its modifiers, but in attributive possessive constructions the head follows its modifiers. Hadza has both verb-initial (VSO) and verb-medial (SVO) order, while all Khoe languages and Sandawe have essentially verb-final (SOV) syntax with modifiers preceding the head. While clause subordination is common in Khoe, coordination involving structures commonly found in verb-serializing languages appears to be widespread in other Khoisan languages. A syntactic characteristic of a number of Khoisan languages is the presence of a multipurpose oblique case marker used to introduce adjuncts, i.e. clausal participants that are not part of the valency of the verb (Güldemann and Vossen 2000:110).

While one language (Nama/Damara) has been described in some detail, the majority of Khoisan languages are virtually unknown, apart from a few grammatical details and/or a wordlist (see Güldemann and Vossen 2000:103). With the exception of Nama/Damara and a handful of other languages, all languages are seriously endangered, many of them on the verge of extinction.

References

Baumann, Hermann (ed.). Die Völker Afrikas und ihre traditionellen Kulturen. Part I, Allgemeiner Teil und südliches Afrika. (Studien zur Kulturkunde, 34.) Wiesbaden: Steiner.

Bleek, Dorothea F. 1956. A Bushman dictionary. (American Oriental Series, 41.) New Haven: American Oriental Society.

Ehret, Christopher. 1986. Proposals on Khoisan reconstruction. SUGIA (Sprache und Geschichte in Afrika) 7(2) 105–30

Greenberg, Joseph H. 1963. The languages of Africa. The Hague: Mouton.

Güldemann, Tom, and Rainer Vossen. 2000. Khoisan. In Heine and Nurse, pp. 99–122.

Heine, Bernd, and Derek Nurse (eds.) 2000. African languages: an introduction. Cambridge: Cambridge University Press.

Honken, Henry. 1977. Submerged features and Proto-Khoisan. In Traill, pp. 145–69.

Köhler, Oswin. 1975. Geschichte und Probleme der Gliederung der Sprachen Afrikas. Von den Anfängen bis zur Gegenwart. In Baumann, pp. 135–373.

———. 1981. Les langues khoisan, In Manessy.

Manessy, Gabriel. 1981. Les langues de l'Afrique subsaharienne. (Les Langues dans le monde ancien et moderne, ed by Jean Perrot.) Paris: Centre National de la Recherche Scientifique.

Meinhof, Carl. 1912. Die Sprachen der Hamiten. Hamburg: Friederichsen.

Sands, Bonny. 1998. Eastern and Southern African Khoisan: evaluating claims of distant linguistic relationships. (Quellen zur Khoisan-Forschung, 14.) Cologne: Köppe.

Schultze, Leonhardt. 1928. Zur Kenntnis des Körpers der Hottentotten und Buschmänner. Jena: Fischer.

Sebeok, T.A. (ed.) 1971. Current trends in linguistics, Vol. 7: Linguistics in sub-Saharan Africa. The Hague: Mouton.

Traill, Anthony (ed.) 1977. Khoisan Linguistic Studies, Vol. 3. (African Studies Institute Communications, Vol. 6.) Johannesburg: University of the Witwatersrand.

———. 1986. Do the Khoi have a place in the San? New data on Khoisan linguistic relationships. SUGIA (Sprache und Geschichte in Afrika) 7(1) 407–30.

Vossen, Rainer. 1997. Die Khoe-Sprachen: Ein Beitrag zur Erforschung der Sprachgeschichte Afrikas. (Quellen zur Khoisan-Forschung, 12.) Cologne: Köppe.

Westphal, Ernst O. J. 1971. The click languages of southern and eastern Africa. In Sebeok, pp. 367–420.

BERND HEINE

Kinship Terms

The inventory of names for family members found in a language can tell us a great deal about the culture of the speakers of that language. How these names are used, both within and outside the family, is also closely related to culture. For these reasons, the study of kinship terms has been important to anthropological linguistics. It has also played a role in theories of word meaning. In cognitive linguistics, kinship terms can provide us with examples of 'unmarked' and 'marked' categories, that is, categories that fit neatly into a symmetrical structure as well as categories that are asymmetrical. This allows us to see the interplay between particular social structures and linguistic universals in forming systems of meanings within a language.

Kinship terminology forms a well-defined subset of vocabulary within a language, and it is a subset that tends to be more structured, symmetrically and hierarchically, than other areas of the lexicon. These characteristics make kinship terms good candidates for a type of meaning analysis known as 'componential analysis'. This type of analysis seeks to break down word meanings into their basic components, typically using a binary classification system. For example, many kinship terms designate the sex of the person referred to, so that *nephew* could be labeled as +male, while *niece* would be labeled as −male. Besides the sex of the referent, other components of kinship terminology are the sex of the speaker, generation (*x* generations older or younger), birth order (used e.g. to distinguish among older or younger siblings), whether the kinship tie is maternal or paternal, whether the relation is by blood or by marriage, and whether the relation is lineal (*father, daughter*) or collateral (*aunt, cousin*). In English, a combination of the components +2 generations, −male, and +lineal would indicate the person called *grandmother*. Different contrasts show up in different languages.

Anthropologists are interested in which contrasts are maintained and which ones are merged or overlap in different societies. In Arapaho, a man would use the term *neyóo* to refer to his brother-in-law, while a woman would use the term *neiθébi* for her brother-in-law. *Neiθébi* actually means something like 'sibling-in-law of a different sex', and it would be used by a man to refer to his sister-in-law. In the Iroquoian languages, there are different words for older and younger sisters and brothers; lineal and collateral relatives are not separated as in English: in Seneca, *noʔyéh* refers to one's mother and her sisters, and *haʔnih* refers to one's father and his brothers. In English, cousins are not distinguished by sex, but in Czech, they are.

The Omaha Indians had a system that has been described as 'skewing' and 'merging' certain categories. Generational distinctions are skewed depending on one's sex, so that a woman's brother and her father are equivalent. In other words, the brother of a female is shifted to an older generation. This leads to the merging of other relations as well: the maternal uncle would be equivalent to the grandfather, and a man's sister would

be equivalent to his daughter. It has been proposed that a system like this one can account for the facts of Proto-Indo-European kinship terminology also (cf. Latin *avus* 'grandfather' and *avunculus* 'maternal uncle').

A look at Proto-Indo-European kinship terms reveals certain features of the social organization of these ancient people. If a woman's brother takes on the status of her father, the men of the family likely held higher status positions of dominance or protection over the women within their own generation. Not only was the culture therefore patriarchal, it was also patrilocal, meaning that when a woman married, she went to live with her husband's family. This is reflected in the greater frequency and number of terms for a woman's in-laws than a man's in the Proto-Indo-European language. Thus, the kinship terms provide us with clues about the role of women in this society.

We can ask whether the categories that are important in distinguishing kinship terms affect the way in which people perceive the world. Does the lack of a particular distinction in the vocabulary of a language mean that people who speak that language are unable to mentally make that distinction? Most would reject this extreme version of linguistic relativity. That is, using the same word for both mother and mother's sisters does not mean that the speaker cannot distinguish who is his/her biological mother. It does, however, imply that one behaves in certain ways toward both the mother and her sisters, perhaps having the same obligations toward them. In this system, as found in the Arapaho language, a man would be obliged to treat his sons and his brother's sons the same way, whenever possible.

As far as behavior is concerned, the extension of kinship terms to refer to people outside the family is also symbolic of a certain expected behavior. In earlier times, African Americans would sometimes refer to any respected elders in the community as *aunt* or *uncle.* Turkish villagers, when speaking to a person of their own generation, can use either the term for mother's brother (*dayi*) to show respect or the term for father's sister's son (*aga*) to express decreased solidarity. Chinese speakers can use terms for elder uncle, younger uncle, or younger aunt to nonfamily members

to denote relative status and a person's attitude of humility toward an older addressee.

Family terms frequently extend metaphorically even into nonhuman realms. Gods and goddesses are called by the terms for father and mother. The Navajo have the same concept that some English speakers do of 'Mother Earth', which they further extend to agricultural fields and corn. Words for brother and sister have been used in various Indo-European and other languages to denote a spiritual kinship as well as a physical one. Whether our behavior toward the earth or toward our fellow human beings can ever parallel behavior expected toward our closest family members remains to be seen.

References

Casson, Ronald. 1981. The semantics of kin term usage: transferred and indirect metaphorical meanings. Language, culture, and cognition, ed. by R. Casson, 230–44. New York: Macmillan.

Friedrich, Paul. 1965. Proto-Indo-European kinship. Ethnology 5. 1–36.

Goodenough, Ward. 1956. Componential analysis and the study of meaning. Language 32. 195–216.

Goodenough, Ward. 1965. Yankee kinship terminology: a problem in componential analysis. Formal semantic analysis, ed. by E. Hammel, 259–87. American Anthropologist 67.

Greenberg, Joseph H. 1980. Universals of kinship terminology: their nature and the problem of their explanation. On linguistic anthropology: essays in honor of Harry Hoijer, ed. by Jacques Maquet. Los Angeles: Undena for UCLA Anthropology Department.

Greenberg, Joseph H. 1966. Language universals. The Hague: Mouton. 9–32.

Heath, Jeffrey, Francesca Merlin, and Alan Rumsey (eds.) 1982. The languages of kinship in aboriginal Australia. Sydney: University of Sydney.

Lounsbury, Floyd. 1964. The formal analysis of Crow- and Omaha-type kinship terminologies. Explorations in Cultural Anthropology, ed. by Ward H, 351–93. Goodenough. New York: McGraw-Hill.

Lounsbury, Floyd. 1964. The structural analysis of kinship semantics. Proceedings of the IXth International Congress of Linguists, ed. by H. Lunt, 1073–93. The Hague: Mouton.

Nogle, Lawrence E. 1974. Method and theory in the semantics and cognition of kinship terminology. The Hague: Mouton.

Schneider, David M. 1980. American kinship: a cultural account, 2nd edition. Chicago: University of Chicago Press.

ELLEN JOHNSON

Kituba

Kituba, a.k.a. Kikongo-Kituba, is a contact-based Bantu language spoken in the western part of the Democratic Republic of Congo—from the Kwilu River all the way to the Atlantic Ocean—and in the southern part of the Republic of the Congo. It is known by several other names, including *Kikongo ya leta* 'Kikongo

of the colonial agent (*leta*)', *Kikongo ya bula-matadi* 'Kikongo of the rock-breaker' (in reference to rock-blasting during the construction of the railroad from the Atlantic coast to Kinshasa), and *Mono kutuba* 'I speak'. Outside the Bakongo area, west of Kinshasa, it is better known by the name *Kikongo*, in reference to the fact that most of its vocabulary comes from one of the Kikongo language varieties associated with the Bakongo people. In the Republic of the Congo, where some of the Bakongo people live, it is called *Ikele ve* 'it isn't', in reference to the nonnative way of negating the verb *kele* 'be'. Only an accident of history has thus favored the name *Kituba* in linguistics.

Although the emergence of Kituba is undeniably associated with the exploitation colonization of the Congo Basin by Belgium (roughly, east of the Congo River) and by France (west of the River), its roots lie in the usage of (Kikongo-) Kimanyanga as a trade language during the precolonial period. Manyanga (situated a few dozen miles west of present-day Kinshasa) was already an important trade center before the exploitation colonization of the Congo Basin—which started soon after Africa was apportioned to European colonial powers in the Berlin Treaty in 1885. During the Kongo Kingdom and its subsequent trade colonization by Europe, several caravans from long distances in the interior went through Manyanga. Kimanyanga, its Kikongo variety, served as the lingua franca. The Belgian and French colonizers adopted it to communicate not only with the Bakongo people in the territory between Kinshasa and the Atlantic coast but also with other indigenous populations farther in the interior. They took it as far east as the Kwilu River. They also brought colonial auxiliaries from West Africa and laborers from as far east as Zanzibar to build the railroad between the Atlantic coast and Kinshasa. (Labor migrations were a common phenomenon during the exploitation colonization of Africa by Europeans during the late nineteenth and early twentieth centuries.) The language contact that ensued from these population movements and contacts produced Kituba from Kimanyanga. There are at least three major dialects of Kituba to date: (1) the Western dialect, spoken in the former Kongo Kingdom south of the Congo River (i.e. in today's Democratic Republic of Congo); (2) the Northern dialect, spoken north of the Western dialect, in the Republic of the Congo; and (3) the Eastern dialect, spoken from the Kwango to the Kwilu rivers in the Democratic Republic of Congo. As explained below, they reflect specific influences of the more indigenous languages on Kituba. This has been facilitated by the fact that it is spoken as a second language in the rural areas, although more and more children have been acquiring it at school because it serves as the medium of education.

Overall, Kituba differs from Kikongo languages and other Bantu languages in the following structural respects: (1) Kituba does not use tone or pitch variations to distinguish meaning to the same extent as the other languages; (2) it has a less complex word structure; and (3) it has lost e.g. the agreement system typical of Bantu languages, in which the inflectional pattern of the subject noun matches that of the verb.

A large proportion of Kituba words have the same intonational system as Swahili. In the case of verbs, the position of the intonational stress changes, depending on whether or not it has a tense-aspect suffix. Thus, *móno kwénda* is 'I go' (in the narrative tense), but *móno kwendáka* is 'I went' (in the past tense). However, there is a significant proportion of words that seem to follow the basic high-and-low tone system of Lingala, e.g. *munoko* 'mouth' (with low tones only), *kímbambala* 'old (machine)' (with a high tone on the first syllable only), and *makilá* 'blood' (with a high tone on the last syllable only). Kituba thus incorporates both tone distinctions and pitch accent, which is not typical of Bantu languages.

Unlike in the Bantu canon, Kituba verbs do not carry pronominal or agreement prefixes, nor do they carry any other preverbal marker for tense or negation. Negation is expressed sentence-finally with an invariant marker *vé*, as in *kwísa vé* 'don't come' (literally, 'come not'). The pronouns derive from the emphatic 'independent' pronouns of Kimanyanga (viz. *móno* 'I/me', *ngé* 'you, singular', *yándi* 'he/him/she/her/it', *béto* 'we/us', *béno* 'you, plural', *báu/bô* 'they/them'), and they occupy the same positions as nouns around the verb, as in *Yándi/Pételo mon-áka ngé/María vé* 'He/She/Peter didn't see you/Mary' (literally, 'X saw Y not'). Aspect and mood are expressed by independent words, as in *yándi ké(l)e kwénda* 'he/she is going' (progressive aspect), *yándi lénda kwénda* 'he/she may/can go' (potential mood), and *yándi ata kwénda* 'he/she will go' (future aspect).

In complex noun phrases, modifiers are connected to the head noun by *na* if they are pronominal and by *ya* in other cases, as in *mukandá na móno* 'my book' (literally, 'book of me'), *mukandá ya Pételo* 'Peter's book' ('book of Peter'), *mukandá ya ngé tang-á(k)a* 'the book that you read' ('book of you read'), and *bíma ya kú-dia* 'food items' ('thing of eat').

As a nonnative lingua franca in rural areas, Kituba is often spoken with different accents that reflect its speakers' linguistic backgrounds. The 'accent' may be identified mostly in word choice (for instance, when a speaker imports a term from his or her ethnic language) and in some grammatical features. For instance, in the western dialect, the structure of the progressive construction uses the connective *na,* as in *móno kéle na kú-dia* 'I am eating' (literally 'I am LOC eat'). The

counterpart of the construction in the eastern dialect is *móno ké(l)e (kú-)dia*, without the connective. Within the eastern dialect, the pattern for commands varies. Up the Kwilu River, *kwis-éte* 'come, please' is the normal imperative, whereas elsewhere the bare imperative *kwísa* 'come' will do and will not offend the addressee. In the same variety, the connective *tí* is used in *yándi kwisáka tí bô* 'he/she came with them', whereas elsewhere in the eastern dialect the same general purpose connective *na* is used.

Whereas varieties spoken in rural areas often reflect influence from the more indigenous ethnic languages, those spoken by the educated reflect French influence. The rural varieties are looked down upon, but their educated counterparts are at best criticized by purists; they are well tolerated. The source language for the vocabulary is no longer considered as the norm from which the other varieties putatively deviate. Rural and educated varieties just deviate from the urban non-Frenchified norm. In fact, French, the official language of both the Democratic Republic of Congo and the Republic of the Congo, is not even part of the Kituba continuum, although it is the source of many lexical borrowings. Relative to Kimanyanga, the main vocabulary source, Kituba seems to have autonomized to an extent seldom reached by most creoles of the Caribbean and the Indian Ocean. Kituba speakers hardly think they speak a dialect of ethnic Kikongo, although they use the latter name in reference to their language in the eastern dialect region. It is interesting that Kituba is hardly ever included in the classification of Bantu languages, which reflects an old linguistic tradition of not including acknowledged contact-based varieties in classifications of language families. Yet Lingala, which has a similar history, is typically included among the Bantu languages.

References

Fehderau, Harold. 1966. The origin and development of Kituba. Ithaca, NY: Cornell University dissertation.

Mufwene, Salikoko S. 1997. Kitúba. Contact languages: a wider perspective, ed. by Sarah G. Thomason, 173–208. Amsterdam: John Benjamins.

Samarin, William. 1989. The black man's burden: African colonial labor on the Congo and Ubangi rivers, 1880–1900. Boulder, CO: Westview Press.

SALIKOKO MUFWENE

Korean

Although the Korean language shares many characteristics with the Japanese and Altaic languages, its genetic relationship is unknown. Its syntactic structure most closely resembles Japanese, but its phonological system differs markedly. Morphologically, it displays a so-called agglutinating characteristic, i.e. words are formed from roots and affixes.

The oldest written records about Korean are in third-century Chinese, eighth-century Japanese '*Nihonshoki*' (Chronicles of Japan, 720) historical records, in which are written Korean place names, and in references to '*Samkwuk-saki*' (Chronicles of the three kingdoms, 1145) and '*Samkwuk-yusa*' (Memorabilia of the three kingdoms, 1285) of the *Goryeo* (or *Kolye* in the Yale Romanization system) dynasty. Korean is only fully recognizable in historical records after 1443, when the *Hangeul* (or *Hangŭl*, *Hankul*) writing system was created, allowing full written representation of the Korean language. Until then, the language was written mainly with Chinese characters and supplementary characters called *Gugyeol* (or *Kugyŏl*, *Kwukyel)*, but these were only able to present a partial picture of the language at the time.

With regard to the historical division of Korean, Lee (1972) calls the period before the *Goryeo* dynasty 'Old Korean', and the *Goryeo* dynasty period 'Early Middle Korean', the period from the end of the fifteenth century in the *Joseon* (or *Chosŏn*, *Yi*) dynasty and up to the Japanese invasion in the late sixteenth century 'Late Middle Korean', the period up to the late nineteenth century 'Modern Korean', and after the nineteenth century 'Contemporary Korean'. During the transition from Middle Korean to Contemporary Korean, some sound changes and various grammatical changes occurred.

Area of Use and Variants

Korean is the national language of the Democratic People's Republic of Korea in the northern part of the Korean Peninsula, and the Republic of Korea in the south. Koreans are an almost completely homogenous ethnic group. North Korea has a population of 22.27 million people, while South Korea's population is 47.27 million people. Additionally, over 1.8 million ethnic Koreans live in China and more than 400,000

reside within the borders of the former Soviet Union. As a result of Stalin's forced settlements, Koreans also inhabit Central Asian countries such as Kazakhstan and Uzbekistan. Between 1910 and 1945, Korea was in a state of annexation by Japan and forced migration to Japan occurred during this era. There are an estimated 700,000 Koreans living in Japan. Korean emigration has also extended its reach to countries like the United States and Australia. Most Korean speakers outside of the Korean Peninsula live in bilingual societies. Now, there are more than 75 million speakers of Korean worldwide.

The national languages of South Korea and North Korea are basically one and the same, but in South Korea the standard language is based on the Seoul dialect, while in North Korea the standard language is based on the Pyeongyang dialect and is called *mwunhwae* 'the language of culture'. Because of this, there are slight variations in vocabulary, pronunciation, and orthography.

Dialect

Korean dialects can be roughly divided into the following six categories. Following the Korean peninsula from north to south, they are: the northeast dialect in Hamgyeong and Ryanggang Provinces, the northwest dialect in Pyeongan and Jagang Provinces, the central dialect in Hwanghae, Gangwon, Gyeonggi and Chungcheong Provinces, the Southeast dialect of Gyeongsang Province, the southwest dialect in Jeolla Province, and Jeju dialect in Jeju Province. The Jeju dialect retains the Middle Korean vowel 'ʌ', which is now lost in Contemporary Korean, and the southwest dialect makes no distinction between the vowels 'ey' and 'ay' or between 'uy' and 'e'. Also, the fortis 'ss' is not present in the southeast dialect. In the northwest dialect, the affricate [tʃ] is pronounced [ts]. Middle Korean had a pitch accent system to distinguish word meanings, but in Contemporary Korean only the northeast and southeast dialects retain a pitch accent, while it has completely disappeared from others such as the central dialect.

Writing System

The Korean writing system, which is today called *Hankul* in South Korea and *wulikul* in North Korea, was created under the guidance of King Sejong, the fourth king of the Joseon dynasty. *Hankul* first appeared on the world stage in written form in 1446 in *Hunminjeongeum* (or *Hunminjŏngŭm*, *Hwunmincengum*), a woodblock publication containing details of the reason for its creation, construction basics, and usage examples. In other words, *Hankul* is unique in the fact that it was carved

and pressed into existence in the form of a book about *Hankul*. In *Hankul*, the written representation form of consonants is based on the forms of the point of articulation. From the following year, 1447, they were used in the publication of the official hymn of the *Joseon* dynasty 'Yongpi-echenka' and the rhyme dictionary 'Tongkwuk-cengwun', and were subsequently used in the publications of many books. Other than book publications, the official script at the time was written with Chinese characters; *Hankul* was mostly used for private communication, but at the end of the nineteenth century and the start of the modern age, *Hankul* came into general use. Today, the literacy rate in both North and South Korea is nearly 100%.

Each Korean letter represents one sound, that is, a letter of an alphabet, which are combined and arranged to form characters. Each combination of letters, or a character, represents one syllable. Therefore, Hankul simultaneously has a single sound, alphabetic characteristics like the Roman alphabet and syllabic characteristics like Chinese characters or Japanese kana:

$$k + a + m > \begin{matrix} k & a \\ & m \end{matrix}$$

ㄱ + ㅏ + ㅁ > 감

Hankul has several methods of transcription: it can be Romanized according to the McCune–Reischauer method, the Yale method, or official methods designated by the governments of North and South Korea. The Yale method is basically used in this writing. The *Hankul*, its Romanized transcriptions, and main pronunciations are shown in Table 1.

Phonetics and Phonology

Words in contemporary Korean are formed from at least one syllable. For most words, the basic vocabulary stem is one or two syllables long. Syllables generally fit this pattern: consonant + semivowel + vowel + consonant. A vowel always forms the core of the syllabic, and one consonant may come on each side of the vowel. For basic vocabulary, a relatively large number of words end with a consonant.

These are the eight vowel phonemes present in the Seoul dialect:

ㅣ i [i]   ㅡ u [ɨ]   ㅜ wu [u]
ㅔ ey [e]
ㅐ ay [ɛ]   ㅗ o [o]
　　　　ㅓ e [ɔ]
ㅏ a [a]
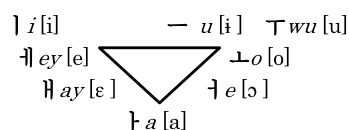
Of the eight vowels, the back vowel *wu* and *o* are rounded vowels, while the others are unrounded. The shape of the lips when pronouncing *i* and *u* are nearly

TABLE 1
Consonants

	IPA	McCune-Reischauer	Yale	North Korean 1955	South Korean 1983	South Korean 2000
ㅁ	m	m	m	m	m	m
ㅂ	p	p/b	p	p	p/b	b/p
ㅃ	ʔp	pp	pp	pp	pp	pp
ㅍ	pʰ	p'	ph	ph	p'	p
ㄴ	n	n	n	n	n	n
ㄷ	t	t/d	t	t	t/d	d/t
ㄸ	ʔt	tt	tt	tt	tt	tt
ㅌ	tʰ	t'	th	th	t'	t
ㄹ	r/l	r/l	l	r	r/l	r/l
ㅅ	s	s	s	s	s,sh	s
ㅆ	ʔs	ss	ss	ss	ss	ss
ㅈ	tʃ	ch/j	c	ts	ch/j	j
ㅉ	ʔtʃ	tch	cc	tss	tch	jj
ㅊ	tʃʰ	ch'	ch	tsh	ch'	ch
ㄱ	k	k/g	k	k	k/g	g/k
ㄲ	ʔk	kk	kk	kk	kk	kk
ㅋ	kʰ	k'	kh	kh	k'	k
ㅎ	h	h	h	h	h	h
ㅇ	n	ng	ng	ng	ng	ng

Vowels

	IPA	McCune-Reischauer	Yale	North Korean 1955	South Korean 1983	South Korean 2000
ㅏ	a	a	a	a	a	a
ㅓ	ɔ	ŏ	e	ŏ	ŏ	eo
ㅗ	o	o	o	o	o	o
ㅜ	u	u	wu	u	u	u
ㅡ	ɨ	ŭ	u	ŭ	ŭ	eu
ㅣ	i	i	i	i	i	i
ㅔ	e	e	ey	e	e	e
ㅐ	ɛ	ae	ay	ai	ae	ae
ㅑ	ja	ya	ya	ya	ya	ya
ㅕ	jɔ	yŏ	ye	yŏ	yŏ	yeo
ㅛ	jo	yo	yo	yo	yo	yo
ㅠ	ju	yu	yu	yu	yu	yu
ㅒ	jɛ	yae	yay	yai	yae	yae
ㅖ	je	ye	yey	ye	ye	ye
ㅘ	wa	wa	wa	wa	wa	wa
ㅙ	wɛ	wae	way	wai	wae	wae
ㅚ	we	oe	oy	oi	oe	oe
ㅝ	wɔ	wŏ	we	wŏ	wo	wo
ㅞ	we	we	wey	we	we	we
ㅟ	wi	wi	wi	wi	wi	wi
ㅢ	ɨi	ŭi	uy	ŭi	ŭi	ui

the same. Also, in the present-day Seoul dialect, distinction of pronunciation has all but disappeared between the mid vowel *ey* and low vowel *ay*. The lack of distinction between these two vowels has even led to the disappearance of the distinction between the subjective form of the first-person pronoun *nay-ka* 'I' and the subjective form of the second-person pronoun *ney-ka* 'you'. In the present-day Seoul dialect, the first

person 'I' is pronounced *ney-ka* and the second person 'you' is pronounced *ni-ka*, or *ne-ka* by the younger generation. In addition to the above-mentioned vowels, the central vowel allophone *e* [əː] was used by members of previous generations, but it has also all but disappeared. *oy* [ø] has been diphthongized to [we]. It is believed that the disappearance of these vowels has progressed dramatically since the 1960s.

A distinction existed between long and short vowels in the first syllable of words, which were used to distinguish meaning such as [nun] (eye) and [nuːn] (snow), but this long vowel is also rapidly disappearing from the Seoul dialect, with most middle-aged and younger speakers pronouncing the long vowel as a short vowel.

There are two semivowels that form a rising diphthong: *j* and *w*. *j* comes before all vowels except *i* and *u*, and *w* comes before all vowels except *wu*, *o*, and *u*.

Consonants can be categorized based on both the point and manner of articulation, as in Table 2.

A distinguishing trait of stops and affricates is that they have three different series. /p, t, c, k/ are called lenis and are nonaspirated, /ph, th, ch, kh/ are voiceless aspirated, and /pp, tt, cc, kk/ are called fortis and are voiceless sounds with tensing of the glottis. The fricatives include the aspirated alveolar sound /s/ and the fortis sound /ss/, as well as the larynx sound /h/. At the beginning of words, /h/ becomes a fricative [h], but in the middle of words it becomes a voiced [ɦ] or is dropped. At the beginning of words, the lenis /p, t, c, k/ become the voiceless [p][t][tʃ][k], and between other voiced sounds they become voiced [b][d][dʒ][g]. The liquid sound /l/ is a flap [ɾ] at the beginning of syllables or before /h/, while after /l/ or at the final position in the syllable it becomes a lateral [l] or a retroflex lateral [ɭ].

One trait of the stops is that when they come at the end of a syllable, the labial [p], lingual [t], and velar [k], instead of being pronounced explosively, are stopped and formed implosively.

When consonants appear next to each other, various kinds of consonant assimilation can occur, such as nasalization.

Grammar

Korean has two modes of existence, as spoken or written language. The written language is not simply a mapping of the spoken language; a certain differentiation of vocabulary and grammar is allowed.

The syntactic structure of a Korean sentence is very similar to that of Japanese. The word order is S(ubject)–O(bject)–V(erb), and modifiers are located in front of the word to be modified (example provided with rough literal and free translations):

> *naˉnun ecekkey alumtawuˉn kkwumˉul kkwuˉessˉta.*
> I-topic yesterday beautiful dreamˉobject dreamˉpast.
> 'I dreamed a beautiful dream last night'.

Sentence types can be divided into predicate sentences integrated by a predicate (verb or adjective), and nonpredicate sentences that are not integrated by a predicate. As a general rule, spoken language has a higher occurrence of nonpredicate sentences than written language. Unlike languages such as English, the subject is a component of the sentence that is only present when required, and the existence of a predicate does not necessarily guarantee that a subject will also be present. Even in written language in predicate sentences such as text in novels, it is normal for there to be no subject in half the cases. Therefore, not only sentences but also clauses may be considered a complete clause if even just a predicate is present.

Predicates take various converb forms to create subordinate clauses such as manner clauses, reason clauses, conditional clauses, and opposite clauses. The elements that can be contained within a subordinate clause, such as topic words, subjects, or adverbials, are largely predetermined by their functions. Generally, the element that makes the proposition is situated in the middle of the sentence, and the modal element forms a hierarchy that surrounds it. Also, as another example of fixed hierarchy structure, whether or not a subordinate clause contains other subordinate clauses depends on its function. Multiple clause structures like the one below are possible, but one cannot switch the relationship between the clauses:

> [[[[*thipi-lul poˉmyense*] *papˉul mekˉuˉmyen*] *kenkang ey an cohˉuˉnikka*] *papˉputhe mekˉeˉla*].
> [[[[while watching TV] eating] is bad for your health] so eat first]. 'Watching TV while eating is bad for your health, so eat first'. [[[[manner clause] conditional clause] reason clause] main clause]

TABLE 2

	Labial	**Alveolar**	**Alveolo-palatal**	**Velar**	**Larynx**
Stops	p,ph,pp	t,th,tt		k,kh,kk	
Fricatives		s,ss			h
Affricates			c,ch,cc		
Nasals	m	n			ng
Liquid			l		

In languages like English, the person and number of the subject dictate the predicate's form, but in Korean, the form the predicate takes is determined by what degree the speaker respects the listener—whether the listener is being treated politely or nonpolitely. This treatment level is called the speech level. For example, the verb *ha·ta* 'to do' originally took these six conjugational forms spanning four different speech levels: *ha·pnita, hay·yo*; *ha·o*; *ha·ney*; and *ha·nta, hay*. But over the past 50 years, the modern Seoul dialect has rapidly evolved to use only the polite forms *ha·pnita* and *hay·yo*, and the nonpolite form *hay*. Of the polite forms, *ha·pnita* is formal and *hay·yo* is informal, but the latter is used far more frequently in daily conversation. The nonpolite *hay* form is used when addressing children or subordinates, or in situations where the speaker is the same age or older than the listener and they have a close relationship. The *ha·o* and *ha·ney* forms are rarely used except in very limited forms of a kind of pseudospeech, such as television dramas. *Ha·nta* is used as the basic speech level form for written language in news, academic, and fiction texts. Furthermore, speech levels in spoken language can undergo dramatic level shifts, even within the same conversation with the same speakers and listeners.

In addition to the speech level paradigm, which is listener-oriented, there is the honorific/nonhonorific paradigm, which is referent-oriented. The honorific form is made by combining the verbal with the suffix *·si·*, as in *ha·nta* (nonhonorific) and *ha·si·nta* (honorific). The honorific form is always used when the speaker expresses an event in reference to an older or superior person(s); usage of the nonhonorific form would be considered rude.

Noun groups are called nominals (*cheyen* in Korean). Nouns, pronouns, and numerals belong to this group. The nominal declension system is formed by agglutinating dependent morphemes such as *salam* (person: nominative), *salam·i* (the person [did something]: subjective), *salam·uy* (person's: genitive), *salam·ul* (a person: accusative), *salam·hanthey* ([someone gave something] to a person: dative), *salam·hantheyse* (from a person: ablative), *salam·ulo* (by a person: instrumental), *salam·a* (oh, person!: vocative), and *salam·hago* (with a person: comitative). Of these, the subjective *·i/ka*, accusative *·ul/·lul*, instrumental *·ulo/·lo*, and vocative *·a/·ya* have multiple forms; the former is used with nominals that have consonant stems, while the latter is used with nominals that have vowel stems. For the subjective and dative cases, as in *sensayngnim·kkeyse* (the teacher being the subject) and *sensayngnim·kkey* (to the teacher), the honorific form differs from the nonhonorific form. Also, the dative case for animate nominals is *salam·hanthey* in spoken language but *salam·eykey* in written language, while inanimate nominals use the dative form *·ey* as in *san·ey* (to the mountain). For the inanimate form, locative·elative forms also exist: *san·eyse* (at the mountain, from the mountain). Special vocative *·iye/·ye* and comitative *·kwa/·wa* written language forms also exist. Additionally, *·uy* is a genitive, which means 'N1's N2' or 'N2 of N1' when used as 'N1·*uy* N2', but only when the genetive is specified; a rudimentary genitive meaning is also understood from the basic construction 'N1 N2' without *·uy*: *wuli nala kkoch* 'we' 'nation' 'flower' > 'our nation's flower'.

There are other markers in addition to those just mentioned, such as the topic or contrast marker *·un/·nun*, or markers that set limitations on meaning such as *·man* 'only' or *·to* 'also, as well'. These markers come after nominals. Also, it is possible to string two or more markers, as in *ku salam·hanthey·man·un* 'only to that person'.

The morphemes that present themselves with nominals are thought to be dependent words and are called particles or postpositions by some scholars, or endings or clitics by others. Regardless, these case and limitation markers are attached outside of the nominative form, and they are not independent words, but they are also different from the endings that occur within words in languages such as Russian or German.

As for word classes, in addition to verbal and nominal, there are also adverb, interjection, and adnominal. Adnominals modify nominals, such as *on* 'the whole', in *on seysang* 'the whole world' or *say* 'new', in *say sensayngnim* 'new teacher', but they themselves do not carry inflection or declension.

Vocabulary

Depending on origin, Korean vocabulary can be divided into three lexical strata: (1) native words, (2) Sino-Korean words, or (3) foreign loanwords. Native words are words considered to have their origin in the Korean language, such as *hana* 'one', *nwun* 'eye', *salam* 'person', *salang* 'love', or *ha·ta* 'to do'. Sino-Korean words are words that have their root in classical Chinese like *il* 'one', *san* 'mountain', or *chayk* 'book'. Foreign loanwords are borrowed words from English or other languages, such as *khemphyuthe* 'computer' or *khephi* 'coffee'. The Sino-Korean category also includes Japanese words made from Chinese characters, such as *chelhak* 'philosophy' or *yepse* 'postcard'. Newspapers and research papers are usually more than 50% Sino-Korean word content, while at the other end of the spectrum, literary works can have more than 70% native word content. Generally speaking, basic Korean vocabulary contains more native words than the other two types. Numerals have both native word and Sino-Korean word forms, and each is used depending on the co-occurring words.

Onomatopoeia is even more plentiful in Korean than in Japanese, with over 8,000 examples found in a typical dictionary. Also, even for a single adjective like 'red', vowel/consonant alternation results in a variety of forms like *ppalkah˘ta*, *ppelkeh˘ta*, *palkah˘ta*, and *pelkeh˘ta*, where the pronunciation shift serves to distinguish between various nuances through a well-developed system of sound symbolism. Korean also has a rich abundance of adjectives.

References

Chang, Suk-Jin. 1996. Korean. Amsterdam and Philadelphia: John Benjamins.

Hankwuk Cengsin˘mwunhwa Yenkwuwen (The Academy of Korean Studies). 1991. Hankwuk Mincok˘mwunhwa Taypaykkwa Sacen (Encyclopedia of Korean Culture). Seongnam, Korea: Hankwuk Cengsin˘mwunhwa Korejskogo.

Kholodovich, A.A. 1954. Ocherk grammatiki korejskogo jazyka (Introduction to Korean Grammar). Moskva: Izdatel'stvo literatury na inostrannykh jazykakh.

King, J.R.P. 1987. An introduction to Soviet Korean. Language research, Vol. 23, No. 2. Seoul: Language Research Institute: Seoul National University.

Kono, Rokuro. 1979-1980. Kono Rokuro Chosakushu (Selected writings of Kono Rokuro). Tokyo: Heibonsha.

Kwuklip˘kwuke˘yenkwuwen (National Academy of Korean Language). 1999. Phyocwun Kwuke-taysacen (Standard Korean dictionary). Seoul: Twusantonga.

Lee, Hyun Bok. 1989. Korean grammar, Oxford: Oxford University.

Lee, Ki˘moon. 1972;1998. Sincengphan Kwukesa Kaysel (Revised introduction to the history of Korean language). Seoul: Thayhaksa.

Martin, S.E., Lee,Y.H, Chang, S-U. 1968. New Korean English dictionary. Seoul: Minjungseorim.

Martin, S.E., Martin, S.E. 1992. A reference grammar of Korean. Tokyo: Charles E.Tuttle.

Nam, Ki˘Shim, and Ko Yong˘kun. 1985;1993. Phyocwun Kwuke Mwunpeplon (Standard Korean grammar). Seoul: Thapchul-phansa.

Noma, Hideki. 2002. Hankwuke Ehwiwa Mwunpep˘uy Sangkwan˘kwuco (Korean: correlative construction of vocabulary and grammar). Seoul: Thayhaksa.

Sohn, Ho˘min. 1994. Korean. London: Routledge.

Suh, Cheong Soo. 1996. Kwuke-mwunpep (Korean grammar). Seoul: Hanyang University Press.

HIDEKI NOMA

See also **Altaic; Japanese**

Krio (and West African Pidgin English)

Krio is an English-lexicon creole spoken natively by about half a million people in the Freetown area of Sierra Leone, but is used as an auxiliary language by virtually all the country's inhabitants. The foundations of the language were laid in 1787, when a settlement was founded in Sierra Leone intended to receive freed Blacks from Britain. These were later joined by two major groups of immigrants. The largest of these was a contingent of about 1,200 American ex-slaves who had fought on the British side against the Americans, and who arrived by way of Nova Scotia. A couple of years later, 550 Jamaican Maroons (slave rebels) were transported to the new colony. As the groups merged, Krio emerged as the language of the new population.

After the abolishment of the slave trade in 1807, the British navy began intercepting slave vessels bound for the Americas, landing the recaptives (in all, almost 100,000) in and around Freetown. Within a short time, they outnumbered the original settlers. Given their ethnolinguistic heterogeneity, the Recaptives also adopted Krio as their language, and by the second half of the nineteeneth century, all Freetown ethnicities had merged to one single Creole population.

Given the settlement history of the Freetown peninsula, it is not surprising that Krio shows similarities with (in this order) both Gullah and Jamaican Maroon speech, as well as with preexisting African varieties of Pidgin English. Although the question is not uncontroversial, most observers agree that Krio is in essence a New World creole.

Compared to other Atlantic English-lexicon creoles, Krio is typologically rather distant from its lexifier, something that is visible not least in its phonology.

Krio has seven oral and five nasal vowels, but it has been suggested that earlier forms of the language operated with a five-vowel system. In the consonantal system, the lack of /z/ and /h/ is noteworthy. Sporadic cases of depalatalization of /ʃ/ may suggest that this too was lacking in older Krio. Although /v/ is phonemic, it has become /b/ in the noninitial position. Furthermore, there is reason to assume that earlier forms of Krio had less liberal phonotactics than English, with a marked preference for CV syllables. On the other hand, contact with African languages has introduced coarticulated stops.

Krio is a tone language opposing high and low tones in a small number of minimal pairs.

Lexically, Krio is marked in part by its continuing contact with local languages (mainly Temne, Sherbro, and Mende), but even more by influence from the

languages of the Recaptives. Thanks to these, Krio features hundreds of items of Yoruba origin.

Characteristic features in Krio that appear to be of Transatlantic origin include the completive *dɔn*, the preposed pluralizer *dɛm*, serial comparatives involving *pas*, existential and locative copula *lib*, habitual *blant*, equative copula *na*, progressive *de*, locative copula *de*, 2pl *una*, relativizer *we*, and habitual *kin*. However, a smaller, but far from negligible, number of items can be demonstrated to represent local African developments. These include the multipurpose preposition *fɔ*, the progressive *lib fɔ*, the modal *fit* 'be able to', and the possessive construction of the type 'John his book'.

While Krio has no official status, there does exist a sizeable body of literature, consisting both of original works and translations. In particular, Sierra Leonean radio makes use of it in its news broadcasts.

The familiarity of Sierra Leoneans with western culture and their ability to speak 'English' favored their employment as middlemen in the British colonies. This was of crucial importance for the linguistic makeup of much of West Africa, where varieties (in part) derived from Krio are referred to as West African Pidgin English (WAPE).

Krio was first implanted on the island of Fernando Poo (now part of Equatorial Guinea) by the British, who founded a naval base there in the early nineteenth century. The native-speaking community is only a couple of thousand strong, but the local variety of Krio has developed into the island's lingua franca. Its position was later strengthened by massive immigration of Nigerians in the mid-twentieth century.

Sierra Leone Creoles were engaged in the British conquest of Nigeria, not only because of their being Anglophile but also because of their Recaptive ancestry having retained links with local Nigerian cultures. At least a third of the Nigerian population is now believed to know WAPE, making it (if L2 speakers are included) the most widely spoken pidgin/creole language in the world. Nativization is attested from the early 1960s, and there are a sizeable number of Nigerians speaking WAPE as their first language.

In Cameroon, WAPE was first implanted by missionaries expelled from Fernando Poo by the Spanish.

It then spread inland, until it came to be spoken in most of southwestern Cameroon. Cameroon is officially divided into an Anglophone and a Francophone zone. While the WAPE of the former is essentially an extension of Nigerian WAPE, the latter area presents some unique features due to the absence of English as a prestige variety. It is more conservative (thus closer to its Krio roots), and there is a noticeable impact of French in the lexicon. Cameroonian WAPE may be spoken by as many as 70% of the country's population, including virtually everybody in its southwestern parts. It appears to have acquired a larger proportion of native speakers than Nigerian WAPE.

In Ghana, WAPE is less widespread than elsewhere, and it does not seem to have any native speakers. It is nevertheless used emblematically in several sectors of Ghanaian society, in particular between men and in milieux such as schools, universities, and in the police corps. It was apparently imported in the late nineteenth or early twentieth century by Sierra Leoneans in British service. Related varieties have also been attested in Togo and Côte d'Ivoire, although the prestige of French has prevented its further spread in these countries.

Since the nineteenth century, finally, a variety of Krio called *Aku* is spoken by a community numbering some 6,000 in the Gambia.

References

Dwyer, David. 1967. An introduction to West African Pidgin English. East Lansing: Produced for the United States Peace Corps by the African Studies Center, Michigan State University.

Faraclas, Nicholas. 1996. Nigerian Pidgin. London, and New York: Routledge.

Féral, Carole de. 1980. Le pidgin-english camerounais: Essai de définition linguistique et sociolinguistique. Nice: Publication de l'Institut d'Études et de Recherches Inerethniques et Interculturelles.

Fyle, Clifford, and Eldred Jones. 1980. A Krio-English dictionary. New York: Oxford University Press/Sierra Leone University Press.

MIKAEL PARKVALL

See also **Nigeria**

Kriol (Roper River Creole)

Kriol is one of two large nontraditional indigenous languages of Australia (the other being Kep Yok (Cape York Creole), spoken in Northern Queensland and the Torres Strait Islands). Kriol is spoken by between 15,000 and 20,000 people in an area covering in excess of 350,000 square kilometers (150,000 square miles). Kriol is spoken

throughout the Kimberley region of Western Australia, in the Northern Territory as far south as the Barkly, and extensively in Western Queensland. The spread of pidgins through this area during the late nineteenth and early twentieth centuries was due to the cattle and livestock industry, but creolization occurred more recently, possibly in mission settlements. There is some evidence that these pidgins actually had their origins in the Sydney area with the earliest European settlements of the late eighteenth century. Kriol is now the primary means of communication among aboriginal people in the above-mentioned area, whose main social mobility is in the direction of broader aboriginal contacts, and not economically 'upward' into the dominant white Australian society.

While early investigations (Kaldor and Malcolm 1979; Sandefur 1979) identified Kriol as a postcreole continuum with Standard Australian English (SAE), Sandefur (1982) makes an excellent case for considering Kriol and SAE to be distinct languages. Apparent mesolects ('lighter' vs. 'heavier' creoles) are actually interlanguages, which Kriol speakers learning SAE use in their dealings with non-Aboriginal people.

Attitudes toward Kriol differ widely throughout the region. Most older speakers do not know the term 'Kriol' at all, and identify their language as either 'Pidgin' or 'English'. Of those speakers who identify it as either 'Kriol' or 'English', attitudes vary from disdain to pride. The attitude of educators had formerly been to treat Kriol as a corrupt or substandard form of English, and to attempt to eradicate it. However, most governments and education systems now recognize Kriol as a language in its own right. In some areas, bilingual Kriol–English programs exist, although these are constantly under threat from both local communities and the (ever-changing) centralized governments. Within some aboriginal communities, the recognition of Kriol is seen as a threat to the maintenance of traditional aboriginal languages, all of which are severely endangered.

Kriol is based on the grammar and vocabulary of SAE, although its phonology, morphology, and to some extent its syntax show enormous influence from traditional aboriginal languages.

Phonology

Most traditional languages lack sounds like *f, v, s, z,* and *th*, and this property is carried over into many varieties of Kriol (many varieties have *s* but not the other sounds). While most traditional languages have three-vowel (a/i/u) systems, Kriol has a five-vowel (a/e/i/o/u) system (which is found among some traditional languages in Northern Australia). Although many traditional languages distinguish long from short vowels, most dialects of Kriol do not. Initial and final consonant clusters are typically either reduced or

made polysyllabic by epenthesis. Some correspondences between SAE and Kriol include:

(1)	SAE		Kriol	
	fellow	[fɛloʊ]	bala/pela	[bala]/[pɛla]
	river	[rivə]	riba	[riba]
	that	[ðæt]	det/jat	[dɛt]/[dʒat]
	visit	[vɪzɪt]	bizit/bijit	[bɪzɪt]/[bɪdʒɪt]
	another	[ənʌðə]	naja	[nadʒa]
	been	[bi:n]	bin	[bɪn]
	all	[ɔl]	ol	[ol]
	sneak	[sni:k]	jinik	[dʒɪnɪk]

Morphology

Pronoun forms come from English morphemes, but the paradigm (given in (2)) is typical of traditional languages. There is a contrast between dual (two people) and plural (more than two), and a contrast between whether the term 'we' includes the person being spoken to or not. There is no gender contrast, and hence no distinction between 'he' and 'she'. There are no separate subject and object forms, unlike both SAE and most traditional languages.

(2)	person	singular	dual	plural
	1 addressee included	mi	yunmi	wi
	addressee excluded	-	mindubala	mibala
	2	yu	yundubala	yubala/yumob
	3	im	dubala	ol/olabat/dei

While Standard Australian English has a plethora of prepositions (e.g. *of, at, for, with*, etc.), traditional languages lack similar separate words, instead using a narrow range of suffixes to indicate these relationships. Kriol does not have such suffixes, and it does not have the broad range of prepositions either. Instead, it utilizes two prepositions, each covering many functions: *la* or *langa* are used to identify locations. *la riba* can mean 'in the river' or 'at the river'. *blanga* marks possession or association, as in *dis ngawu blanga Jingili*, meaning 'this (is the) home of the Jingili people'.

Kriol has a particle *na* that is used to indicate emphasis. *Im bin bogi-bogi* means 'she or he was swimming' (literally: 'she or he been swim-swim'; the verb *bogi* is 'reduplicated' to indicate an action in progress or a repeated action). *Im na bin bogi-bogi*, on the other hand, means 'she or he was swimming' or 'It was her who was swimming'.

If verbs take an object, they are marked with a suffix *-im* in Kriol. *-im* is derived from English 'him'. *Sili baga im bin dringk-im* means 'the silly bugger drank it' (literally 'silly bugger it been drink-it'), and *ol bin pik-im-ap manggo* means 'they picked up mangos' (literally 'they been pick-it-up mango'). Notice that the *im* 'it'

in the first sentence can be left out, since the *-im* marking on the verb already indicates that the verb is transitive, i.e. the sentence is about the fact that the silly bugger drank *something in particular*. Kriol does not make a distinction between definite 'the silly bugger' and indefinite 'a silly bugger'. The context must convey what is meant. Kriol does not distinguish between singular 'mango' and plural 'mangos' either. Here, again, the listener has to decide which interpretation is more appropriate in the situation.

Here is a brief Kriol text from Newcastle Waters in the Barkly Tableland, which illustrates many of the properties discussed above:

> *'Wel, dat dem pipul bin gu solja, wantu fait langa naja pleis, blanga naja pipul, dei bin blanga kool blad. Dey bin jinik ap an fait langa alabala waj deya wel bin fait. Sambala dei bin lusim pipul dei bin get supia tru.'*

Well, some people went off to make war [go soldiers], to fight in another country, another people's country, they went in cold blood. They snuck up and fought with all the people there and there was a battle. Some of them lost people when they got speared through.

References

Harris, John W. 1986. Northern Territory Pidgins and the origin of Kriol. Canberra: Pacific Linguistics C 89.

Kaldor, Susan, and Ian Malcolm. 1979. The language of the school and the language of the Western Australian aboriginal schoolchild—implications for education. Aborigines of the West: their past and present, ed. by Ronald M. Berndt and Catherine H. Berndt, 406–37. Nedlands: University of Western Australia Press.

Koch, Harold. 2000. The role of Australian Aboriginal languages in the formation of Australian Pidgin grammar: transitive verbs and adjectives. Processes of language contact: case studies from Australia and the Pacific, ed. by Jeff Siegel. Montreal: Fidas.

Sandefur, John. 1979. An Australian Creole in the Northern Territory: a description of Ngukurr-Bamyili dialects, Part 1. Work Papers of SIL-AAB, series B, Vol. 3.

——— 1981. Kriol—an Aboriginal language. Hemisphere 25. 252–6.

——— 1982. Kriol and the question of decreolisation. International Journal of the Sociology of Language 36. 5–13.

——— 1991. A sketch of the grammar of Kriol. Language in Australia, ed. by Suzanne Romaine, 204–12. Cambridge: Cambridge University Press .

ROB PENSALFINI

L

Labov, William

Although William Labov took many years to follow his true calling in linguistics, he made good use of the time and the life experiences that he acquired in the nonacademic world. Following ideas propagated by his mentor, Uriel Weinreich, he brought to the field of linguistics the vision that data could be naturalistically collected and empirically tested and analyzed. In the 1960s and 1970s, this was considered an extreme contrast from the ideal speaker/hearer dyad data used by Noam Chomsky and his disciples. Labov elaborated and finessed a style of linguistics that was practically unheard of at the time—sociolinguistics. Using the concepts of language variation, the linguistic variable, and the interactions of social factors with both of these, Labov literally created a new style of linguistics. Using data from Martha's Vineyard, New York City, and Philadelphia, he proved that language change can be studied and documented as it occurs.

After working as a chemist for 11 years, Labov went back to school, and created a *coup d'état* with his master's thesis work. He conducted research on the small island of Martha's Vineyard, Maine. His work on the pronunciations of the vowels of the native islanders was groundbreaking because he found that, although the variation of the vowels seemed random, it was in fact related to occupation, geography, and ethnicity. This was one of the first studies that incorporated social factors into linguistic variation, a style of investigation that began to be referred to as 'variationist' research. This study and further works by Labov are so well accepted as the linchpins of variationist research that the concepts contained in them are often simply referred to as the 'labovian framework', with no capitalization or any attribution as to the year or the research involved. Implicit in the variationist approach is the concept of the 'speech community'. A speech community traditionally referred to a group of people who were delimited solely by linguistic criteria. The definition of a speech community in the variationist approach, however, puts much less emphasis on language as a necessary and sufficient condition for group membership. Instead, it suggests that the three key requirements for membership in a linguistic/social group are the existence of a social community, that patterns of language use are shared by its members, and that norms for the social evaluation of speech are shared by community members.

This shared set of social attitudes posited for a speech community were proposed in what has come to be known as the classic variationist research, Labov's (1966) study of New York City speech. In this work, he established a range of speech styles that included word lists, minimal pairs, reading passages, interview style, and casual style. He posited that the linguistic variable that was under scrutiny (r-fulness or r-lessness, for example) would reveal *patterned* variation by speakers, which could be correlated with the independent variable of stylistic range. His hypothesis was that, since the amount of attention paid to speech is highest (more closely monitored, in Labov's terms) while reading lists of words, and the least amount of attention is paid to pronunciation in a casual setting, there would be a range of variation on going from the most 'standard-like' pronunciation in the most highly monitored speech (the word lists) to

the least 'standard-like' pronunciation in the least monitored speech (the casual style). Although the data collection methods were similar for both the Martha's Vineyard and the New York City studies, they showed diametrically different results. In Martha's Vineyard, the locals were using the 'marked' vowel pronunciation as a way of showing pride in their heritage, while in New York, the 'lower class' pronunciation (r-lessness) gave way to the more prestigious variety of the pronounced 'r' when social factors came into play.

In *Sociolinguistic patterns*, Labov (1972) set out to prove yet again that language should not be separated from the culture and society in which it finds itself. Labov was certainly not the first linguist to advocate this concept. Older works by Meillet (1921) and Fischer (1958) also recognized the fact that social factors and language use are inextricably intertwined and suggested that one could not be adequately described without accounting for the influence of the other. However, Labov carried this idea a step further by applying it not just to synchronic processes of language change, but to diachronic processes as well. He used data from a series of studies done on 'change in progress' to show that various functions of language can determine and force language change, that grammar rules can (to a certain extent) be affected by changes in society, and that linguistic evolution or change is not dysfunctional, and therefore '(not a) massive testimony to original sin' (p. 323).

Labov (1972) also examined the philosophy and views of various linguists about accepting and using social factors as one explanation of linguistic change. He suggested the following dichotomy: a 'social' group that believes that one can see change in progress, follows the 'wave model' of linguistic evolution, and believes that social factors can and do affect language, vs. an 'asocial' group that disregards the effects of linguistic diversity and contact between languages, works with a *Stammbaum* or family tree model of linguistic evolution, and does not believe that social factors can help explain linguistic change. Labov places himself squarely in the first group, and argues persuasively that social variation does have a place in not only synchronic but also diachronic description of change. These musings on philosophy and its interconnection with language are carried even further in Labov (1994), in which he compares linguistic change to geographical change, and notes that both can be either gradual or catastrophic, and that both may put linguists in a better position to correlate the social and linguistic factors and respond better to Meillet's suggestion that social change is the only possible alternative to explain linguistic change. For Labov, variation is a reflex of change in language.

It occurs both throughout time and in the present, and can be shown by studying language change in 'real time', as he did in his Martha's Vineyard study (Labov 1972), or by studying language change in 'apparent time', as he did with his study of the evolution of New York City vowels (Labov 1966, 1972).

In an attempt to quash criticism from those who claim that the variationist approach is too subjective, statistical programs that take into account and correlate social and linguistic variables have been written. This results in quantitatively analyzed data collection and analysis and allows for cross-linguistic and cross-cultural comparisons. Labov has recently written a program called Plotnik for plotting, analyzing, and comparing vowel systems. In addition, in recent years, he has also tackled the issue of forensic linguistics by giving expert testimony in court cases, promoted the cause of accepting African American Vernacular English as an accepted and normal language in and of itself, promulgated the idea that linguists should give something back to the community from which they take their data, and overseen the creation of the 'Phonological Map of North America'.

The most important concept to come out of his research was the idea of the linguistic variable. Although dialectologists have long acknowledged the fact that different people speak in different ways, no one had taken into account the fact that the same person speaks differently on different occasions. It was commonly thought that this intraspeaker variation was random, and therefore unquantifiable. With the advent of variationist sociolinguistics, it can now be seen that the variation is systematic, and is the result of the interaction between linguistic variables and social factors. Rather than variables (or linguistic elements) being in free variation (or random), they suggest that variables have been and are conditioned by certain social and cultural factors. Chambers (1995) puts this quite strongly by noting that correlating linguistic variation as the dependent variable with independent variables such as linguistic environment, style, or social categories is the primary empirical task of sociolinguistics. The linguistic fallout from this belief has resulted in an explosion of work dealing with the effects of gender on language, social class on language, race and ethnicity on language, education level on language, and geographical location on language. The interplay of these factors and their effect on speech can now be used to view language change in progress. In the past, language change had to be tediously reconstructed using ancient texts and sheer speculation. One of the most exciting facets of the variationist approach to linguistics is the fact that it can be used

as a key to the past. Language change is invariant; all languages change. The modern-day methods pioneered by Labov and others can be used as a template from which to examine previous changes that languages have undergone as well as to document the changes that speakers impose on their languages every day.

Biography

William Labor was born in Rutherford, New Jerseyon, December 4, 1927. He recived his B.A. (1948) from Harvard College; M.A. (1963) from Columbia University for a thesis on social aspects of sound change on Martha's Vineyard, tutored by Uriel Weinreich; and Ph.D. from Columbia for work on class differences in the dialects of New York City. He was Assistant Professor, Columbia University in: 1964–1970, and Associate Professor, Professor, University of Pennsylvania (1971–present). He was a Guggenheim Fellow (1970–1971 and 1987–1988), President, Linguistic Society of America (1979), and Fellow, American Association for the Advancement of Science (1997–present). He received honorary Ph.D.s form University of Uppsala (1985), University of Liege (1990), University of York (1998), and Haifa University (1999). He was Sapir Professor, Linguistic Institute in 1986, and is since 1988 to date Editor of *Language Variation and Change* (1988–present).

References

Bailey, Charles-James N.1973. Variation and linguistic theory. Arlington, VA: Center for Applied Linguistics.
Chambers, Jack. 1995. Sociolinguistic theory. Oxford: Blackwell Press.
Fasold, Ralph. 1990. The sociolinguistics of language. Cambridge: Blackwell.
Fischer, John L. 1958. Social influences on the choice of a linguistic variant. Word 14. 47–56.
Labov, William. 1963. The social motivation of a sound change, Word 19.
———. 1966. The social stratification of English in New York City.
———. 1969. Contraction, deletion, and inherent variability of the English copula. Language 45. 715–62.
———. 1972. Language in the Inner City: studies in the Black English vernacular. Philadelphia, University of Pennsylvania Press.
———. 1972 Sociolinguistic patterns. Philadelphia, PA: University of Pennsylvania Press.
———. 1975. What is a linguistic fact. Lisse, The Peter de Ridder Press.
———. 1994. Principles of linguistic change: Vol. 1, Internal factors. Oxford: Basil Blackwell.
Meillet, Antoine. 1921. Linguistique historiique et linguistique generale. Paris: La Societe Linguistique de Paris
Saussure, Ferdinand de. 1916. Cours de linguistique generale. Paris: Payot.
Wardhaugh, Ronald. 1986. An introduction to sociolinguistics. Oxford: Blackwell.
Weinreich, Uriel. 1968. Languages in contact: findings and problems. The Hague: Mouton.
Wolfram, Walt, and Natalie Schilling-Estes. 1998. American English. Oxford: Blackwell.

MEGAN E. MELANÇON

Ladefoged, Peter

Like his teacher David Abercrombie, Peter Ladefoged continues the tradition of British phonetics and linguistics, which includes the names of such scholars as Henry Sweet, Daniel Jones, and John Rupert Firth (all three of whom are featured in the photograph opposite the title page of Asher and Henderson (1981), one of the major sources in phonetic historiography, and with the exception of Firth, specifically credited for their pioneering research by Ladefoged in *A course in phonetics* (2001:ix)). Ladefoged began his career at the age of 17 studying physics, chemistry, and mathematics at Cambridge University, but after a stint in the army, he decided to study poetry. By chance, he happened to run into a man named David Abercrombie, who had just started up a Phonetics Department, and

who was himself the son of a poet, Lascelles Abercrombie. I think it is this early exposure to experimental sciences that propelled Ladefoged to concentrate on experimental phonetics. Ladefoged's student at UCLA, John J. Ohala of the University of California, Berkeley, has called the *maestro* the 'world's leading phonetician', with which I am in firm agreement. Born in England and educated in Scotland, the long-time UCLA professor of phonetics and director of the UCLA Phonetics Laboratory (currently UCLA Research Linguist and Professor of Phonetics Emeritus) is practically a household name in many linguistics departments in the United States and around the world. Countless numbers of students (including mine, ever since the first edition in 1975) have been

introduced to phonetics from the four editions of his *A course*, not to mention his numerous other books and articles (see under *References* for typical examples). Many more readers will be exposed to phonetics from another survey work of his that appeared recently, *Vowels and consonants: An introduction to the sounds of languages* (Oxford: Blackwell, 2001), which is less technical than *A course* in that it omits many detailed phonetic concepts and is couched in simpler language; however, there is, not surprisingly, some overlap between the two, such as with the vowels of English ((pp. 26-30) and *A course* (p. 29, p. 81)) and the contrasting stops in Malayalam ((p. 149)) and *A course* ((p. 141)). At first, however, it may seem more complex than *A course* since it starts with acoustic phonetics (it is Chapter 8 of 11 chapters in *A course*) — a Ladefogian specialty ever since the publication of his *The nature of vowel quality* (1962) and his *Elements of acoustic phonetics* (1962).

One of Ladefoged's most outstanding contributions to phonetic science has been the publication of his phonetic fieldwork and the insights gleaned therefrom. He has conducted field research in Africa, India, China, Papua New Guinea, Australia, etc., recording hundreds of languages over the years abroad or at the UCLA Phonetics Laboratory. These recordings, which are housed in the UCLA phonetics archives, provide the requisite data for the phonetic theory to which Ladefoged has contributed so much (see e.g. Ladefoged 1965). As a pioneer in experimental phonetics (see e.g. Ladefoged 1967), including still and cine x-ray photography (see as far back as Ladefoged 1958), palatography, spectrograms, waveforms, and aerodynamic data techniques, Ladefoged's career has been devoted to the phonetic phenomena of the world's languages (see e.g. Gordon et al. 2000), and as a result of this type of work, there now exists a scientific basis for distinguishing linguistic from nonlinguistic sounds. Thus, through the theoretical and experimental sides of his research (and that of many of his colleagues), we now know enough of the sounds of the world's languages for there to be a book with that very title (Ladefoged and Maddieson (1996)).

For Ladefoged and his student John J. Ohala, phonetics is but part of phonology, the goal of which is to understand all aspects of sound patterning in languages (see Ohala 1994:3053). It is precisely this kind of emphasis that prompts Ladefoged and Maddieson (1996:7) to assert that the typical five-vowel system for many languages is the more distinct set /i e a o u/ rather than their rounded (or more marked) counterparts /y, ø, a, ɣ, ɯ/ (/a/ is neutral and is included in both sets). Further, Ladefoged has demonstrated repeatedly that some phonological statements may be better formulated and understood by reference to the acoustic rather than articulatory properties of sounds. Also, he has shown that vowels are better described in terms of auditory rather than acoustic or articulatory parameters (Johnson et al. 1993).

In addition to his fieldwork, Ladefoged has been influential in linguistic terminology and theory. Ladefoged (1964) e.g. coined the term 'approximant', which has gained widespread acceptance in the profession as synonymous with semivowel and frictionless continuant. It is so named because one articulator *approximates* the position of another. He has demonstrated a number of interesting phonetic laws or tendencies; e.g. that there is no clear one-to-one correlation between syllables and chest-pulse movements, or that apical trills are more common than laminal trills. However, it was as the President of the International Phonetic Association that he seized the opportunity to propound the Kiel Convention of 1989, which was organized to revise the International Phonetic Alphabet (IPA). As a researcher of rare speech sounds himself, such as bilabial trills, voiceless implosives, and velar laterals, he was instrumental in making contributions to the adopted IPA revision. It is safe to report that he was the one who first noticed a flapped labial stop in an ideophonic word in Margi (a Chadic language of Nigeria). Ladefoged should also be credited for inventing new ways of looking at speech sounds. In fact, he is largely responsible for insisting that the IPA chart be regarded as a one-page theory of phonetics.

Distinctive feature theory is yet another area where Ladefoged's work has been peremptory. Critical of the features proposed in Chomsky and Halle (1968), Ladefoged (1971) proposed a new system, which recognized the importance of an auditory basis for some features. As a result of this, it is now accepted that some sounds are to be grouped together on the basis of auditory phenomena. In view of his numerous accomplishments described above, there can be little doubt that he has earned his noteworthy international reputation.

Biography

Peter Ladefoged was born in Sutton, Surrey, England on September 17 1925. He attended Haileybury College, Hertford (1938–1943). He received his M.A. (1951) and Ph.D. (1959) from the University of Edinburgh for a dissertation on the acoustics of vowel quality, with Daniel Jones, recently retired from the chair of phonetics at University College, London, serving as a consultant to the project. He was Assistant Lecturer in Phonetics, University of Edinburgh in 1953–1955 and Lecturer in Phonetics, University of Edinburgh, 1955–1959, University of Ibadan, 1959–1960, and University of Edinburgh, 1960–1961.

He was a Field Fellow in the West African Languages Survey in 1961–1962. He was also Assistant Professor of Phonetics, UCLA, 1962—1963; Associate Professor of Phonetics, UCLA, 1963–1965; and Professor of Phonetics, UCLA, 1965–1991. He: was Chair, Department. of Linguistics, UCLA in 1977–1980, and Fellow, Acoustical Society of America and American Speech and Hearing Association. He was a recepient of the Distinguished Teaching Award and Research Lecturer, UCLA, in 1972 and 1989, respectively. Ladefoged was President, Linguistic Society of America in 1978; President, International Phonetic Association in 1987–1991; Fellow of the American Academy of Arts and Sciences, 1990; and UCLA College of Letters and Science Faculty Research Lecturer in 1991. He received a Gold Medal at the XII International Congress of Phonetic Sciences in 1991. He was also Corresponding Fellow of the British Academy in 1992. He also received an honorary D. Litt. from the University of Edinburgh in 1993, and a Silver Medal from the Acoustical Society of America in 1994.

References

Asher, R. E. and J. A. Henderson (eds.) 1981. Towards a history of phonetics. Edinburgh: Edinburgh University Press.

Chomsky, Noam and Morris Halle. 1968. The sound pattern of English. New York: Harper & Row.

Fromkin, Victoria A. 1994. Peter Ladefoged (1925-). The encyclopedia of language and linguistics, Vol. 4, ed. by R. E. Asher, 1891–2. Oxford: Pergamon Press.

Johnson, Keith, Peter Ladefoged, and Mona Lindau. 1993. Individual differences in vowel production. Journal of the Acoustical Society of America. 94. 701–14.

Ladefoged, Peter. 1958. Use of palatography. Journal of Speech and Hearing Disorders 22. 764–74.

———. 1962. Elements of acoustic phonetics (2nd edition with added chapter on computational phonetics, 1996). Chicago: University of Chicago Press.

———. 1962. The nature of vowel quality. Monograph supplement to Revista do Labóratorio de Fonética Experimental de Faculdade de letras da Universidade de Coimbra.

———. 1964. A phonetic study of West African languages. Cambridge: Cambridge University Press.

———. 1965. The nature of general phonetic theories, Georgetown University Monographs on Language and Linguistics. 18. 27–42.

———. 1967. Three areas of experimental phonetics. Oxford: Oxford University Press.

———. 1971. Preliminaries to linguistic phonetics. Chicago: University of Chicago Press.

———. 1975. A course in phonetics, 4th edition, 2001. Fort Worth: Harcourt.

———. 2001. Vowels and consonants: an introduction to the sounds of languages. Oxford: Blackwell.

Ladefoged, Peter, and Ian Maddieson. 1996. The sounds of the World's languages. Oxford: Blackwell.

Ladefoged, Peter, and D. Everett. 1996. The Status of Phonetic Rarities. Language. 72. 794–800.

Ladefoged, Peter, and Matthew Gordon and Pamela Munro. 2000. Some phonetic structures of Chickasaw. Anthropological Linguistics 42. 366–400.

Laver, John. 1994. Principles of phonetics, Cambridge: Cambridge University Press.

Ohala, John J. 1994. Phonetics, ed. by R. E. Asher, The encyclopedia of language and linguistics, Vol. 6, 3051–3. Oxford: Perganon Press.

ALAN S. KAYE

Lakoff, George

George Lakoff had his first contact with Linguistics while he was an undergraduate student at MIT. There, he worked on the structure of discourse, and tried to put a narrative structure into a transformational framework. In those days, one of the basic tenets of transformational grammar was the independence of syntax from semantics. However, beginning in 1963, Lakoff started to find evidence that this central claim was not totally correct. Together with other linguists such as James D. McCawley, Paul Postal, and John Robert Ross, he developed the theory called Generative Semantics. Its main hypothesis is that there is no principled difference between syntactic and semantic processes. Along with this, there are three other main assumptions: (i) nonexistence of a purely syntactic level of 'deep structure', (ii) the universal-base hypothesis: initial representations of derivations are cross-linguistically identical logical representations, and (iii) the derivation of a sentence is a direct transformational mapping from semantics to surface structure.

Despite its initial success, Generative Semantics started to fade away in the mid-1970s. One of the main reasons was the growing body of evidence about the embodiment of mind from the fields of cognitive science and neuroscience. Work by Charles Fillmore in frame semantics, Paul Kay in color categorization, Leonard Talmy in universal spatial cognition, and Eleanor Rosch in basic categories contradicted one of the major tenets of this theory, namely the belief in formal logic and model-theoretic semantics. This new

research showed that it was not possible to define meaning in terms of reference and truth conditions, in terms of the objective structure of the external world, with no reference to the brain or mind. Therefore, formal logic could not be an adequate theory of natural language semantics, and neither could be transformational grammar of syntax. As a consequence, Lakoff stopped doing Generative Semantics and set about developing a field of Cognitive Linguistics.

Cognitive Linguistics views linguistic knowledge as part of general cognition and thinking; linguistic behavior is not separated from other general cognitive abilities that allow mental processes of reasoning, memory, attention or learning, but is understood as an integrated part of it. One of Lakoff's major contributions to the theoretical foundations of this framework is the theory of conceptual metaphor.

In the late 1970s early 1980s, Lakoff, then working in collaboration with Mark Johnson, discovered that the classical view of metaphor as a mere figure of speech based on inherent similarity to indicate a comparison was empirically false. Instead, they showed that metaphor is part of the cognitive unconscious. It is a cognitive mechanism that allows us to experientially ground conceptual systems and to reason about abstract domains (e.g. *love*) in terms of more concrete domains (e.g. *journeys*). In other words, metaphor is a cross-domain conceptual mapping. Lakoff has applied the metaphor theory to the analysis of the conceptual structure of several disparate disciplines such as poetry, visual arts, politics and social issues, philosophy, and mathematics. His work proves that a system of metaphorical mappings allows the abstract ideas of any discipline to be conceptualized in terms of embodied experience.

In the 1990s, together with other researchers such as Jerome Feldman, Lakoff founded the Neural Theory of Language, which seeks to answer the question of how the physical brain can give rise to concepts and language. The basic foundations of this research program, previously known as the L_0 Group, are drawn from many different disciplines: neurochemistry, neuroscience, neural computation, psychology, and cognitive linguistics. The result is a completely different understanding of language from that in traditional Chomskyan theories. In this framework, language is characterized as: 'embodied' —structured by our bodily and social experience of the world, as well as by the general properties of neural systems and the neural structure of systems such as vision and motor control; 'content-dependent' —physical and biological constraints; 'cognition-dependent' —other aspects of human cognition; 'evolutionary' —use of earlier evolutionary development; and 'learned' —on the basis of innate structure that is not specifically linguistic. In his career as a linguist and cognitive scientist, George Lakoff has been a founder of three different intellectual movements: Generative Semantics, Cognitive Linguistics, and the Neural Theory of Language.

Biography

George Lakoff was Born in Bayonne, New Jersey on, May 24, 1941. The graduated from MIT in 1962 in English Literature and Mathematics, and studied Linguistics with Roman Jakobson, Morris Halle, and Noam Chomsky. He received his doctorate in Linguistics in 1966 from Indiana University for his work on the irregularity in syntax and transformational grammar. He was a Research Fellow and Lecturer in Linguistics at Harvard (1965–1969). Then, he worked as associated Professor of Linguistics at the University of Michigan (1969–1972). He moved to the University of California at Berkeley, where he is Professor of Linguistics since 1972. He has been a fellow of the Institute for Advanced Study in the Behavioural Sciences at Stanford (1971–1972), and a visiting professor at the École des Hautes Études en Sciences Sociales in Paris (1995). He served on the Governing Board of the Cognitive Science Society (1990–1996), as President of the International Cognitive Linguistics Association (1990–1994), and as a member of the Science Board of the Santa Fe Institute (1995–present). Lakoff was a founding member of two recently formed political think tanks: The Rockridge Institute (1998), where he is a Senior Fellow, and the Frameworks Institute (1999), where he is a member of the Advisory Board. He was a recipient of the 2001 Hatfield Scholar Award for socially relevant research by a social scientist, Mark O. Hatfield School of Government at Portland State University.

References

Gibbs Jr., Raymod W., and Gerard J. Steen (ed.). 1999. Metaphor in cognitive linguistics. Amsterdam and Philadelphia: John Benjamins.

Johnson, Mark. 1987. The body in the mind. The bodily basis of language, reason and thought. Chicago: Chicago University Press.

McCawley, James. 1976. Grammar and meaning. New York: Academic Press.

Lakoff, George 1963. Toward generative semantics. MIT Mechanical Translation Project Report.

———. 1970. Irregularity in syntax. Austin, TX: Holt, Rinehart and Winston.

———. 1970. Global rules. Language.

———. 1972. Hedges: a study in meaning criteria and the logic of fuzzy concepts. Papers from the Eight Regional Meeting of the Chicago Linguistic Society.

———. 1974. Berkeley studies in syntax and semantics, Vol. I, ed. by Charles Fillmore and Robin Lakoff.

Lackoff, George, and Mark Johnson. 1980. Metaphors we live by.

———. 1987. Women, fire, and dangerous things: what categories reveal about the mind.

Lackof, George, and Mark Turner. 1989. More than cool reason: a guide to poetic metaphor.

———. 1993. The contemporary theory of metaphor. Metaphor and thought, 2nd edition. ed. by Andrew Ortony

———. 1996. Moral politics: what conservatives know that liberals don't.

———. L_0 — The first five years of an automated language acquisition project. AI Review, Special Issue on Integration of Nautral Language and Vision Processing.

Lakoff, George, and Jerome Feldman, David Bailey, Srini Narayanan, Terry Regier, and Andreas Stolcke. 1996.

Lakoff, George, and Mark Johnson. 1999. Philosophy in the flesh: the embodied mind and its challenge to western thought.

Lakoff, George, and Rafael Núñez. 2000. Where mathematics comes from: how the embodied mind brings mathematics into being.

Ortony, Andrew, (ed.), 1986. Metaphor and thought. Cambridge: Cambridge University Press, 2nd edition. 1993

Regier, Terry. 1996. The human semantic potential. Spatial language and constrained connectionism. Cambridege, MA: MIT Press.

Ruiz de Mendoza, Francisco José. 1997. An interview with George Lakoff. Cuadernos de Filología Inglesa 6(2).

Ungerer, F., and H.-I. Schmid. 1997. An introduction to cognitive linguistics. London: Longman.

IRAIDE IBARRETXE-ANTUNANO

See also **Fillmore, Charles; McCawley, James D.; Metaphor**

Langacker, Ronald

Ronald Langacker had his first contact with Linguistics at the University of Illinois, where he received his training and pursued his early research in generative syntactic theory. His dissertation was a transformational description of aspects of French syntax. The first ten years of his professional career, 1966–1976, were devoted to the study of comparative grammar and historical reconstruction of the Uto-Aztecan language family of Native American languages.

In 1976, he decided that current theoretical approaches to language such as Generative linguistics no longer offered satisfying answers for understanding language. Consequently, he started developing a new linguistic framework called Cognitive Grammar, initially known as Space Grammar, which presents a radically different conceptualization of language and its structure.

In contrast to modular approaches, this new theory proposes that language is an integral part of human cognition and that grammatical structure can be characterized in terms of cognitive processing. Cognitive grammar assumes that (i) grammar is not autonomous with respect to semantics, (ii) grammar can be fully described in terms of 'symbolic units' i.e. form-meaning pairings, and (iii) grammar together with the lexicon (vocabulary), morphology (word-structure and word-formation rules), and syntax form a continuum of symbolic relationships.

This view of grammar as symbolic in nature relies on a very specific understanding of linguistic semantics. In this approach, meaning is not identified with truth conditions (as it is in classical logic), but it resides in 'conceptualization' and 'mental experience'. This includes both fixed and novel conceptions, sensory and motor experience, and a full grasping of physical, social, cultural, and linguistic context. As a consequence, the boundaries between the traditional distinction between linguistic and encyclopedic knowledge are blurred. In conceptualist semantics, the meaning of a lexical item is not restricted to the meaning of the expression designating it, but is open to all the knowledge one may have about that entity itself.

Every expression and every symbolic unit has two equally significant sides for linguistic meaning: the conceptual 'content' it evokes and the particular 'construal' imposed on that content. The content of an expression comprises a set of cognitive domains, each of which corresponds to a different aspect of the expression's semantic value. Therefore, the starting point for semantic description is not a set of semantic features or conceptual primitives, but an appropriate array of integrated conceptions organized at different levels of conceptualization. Equally important for an expression is to know how it is construed. Langacker proposed several different cognitive operations for the structuring of knowledge, different dimensions of construal that a speaker may take into account when describing a certain situation. Among the most significant, we can mention the following: (i) 'specificity', the perception of objects and situations at different levels of detail or granularity; (ii) 'scope', all the possible aspects that a given expression may evoke in its characterization; (iii) 'background' assumptions and

expectations, that is, the distinction between given and new information; (iv) 'perspective', a speaker's viewpoint on the situation, e.g. orientation, vantage point, directionality; and (v) 'prominence', which includes very influential concepts such as profile/base organization, trajector/landmark alignment.

In Cognitive Grammar, an expression's grammatical category depends on the nature of its profile. For Langacker, there are 'things' that profile nominal expressions such as nouns *(dog, table...)*, pronouns *(I, you...)*, determiners *(the, this...)* and even noun phrases *(the large table...)*, and 'relationships' that profile relational expressions; some are atemporal relations such as adjectives *(green, big...)*, adverbs *(well, quickly...)*, and prepositions *(behind, of...)*, and some others are processes or temporal relations such as verbs *(take, sleep...)*. Grammar consists of patterns that combine simpler symbolic structures with symbolic structures of greater complexity, the latter called 'grammatical constructions'.

Another important feature of this model is that it is a usage-based approach. That is to say, it is based on how speakers actually speak and understand language, which, in turn, is composed of networks of units that become conventionalized or entrenched through usage.

The first general description of Cognitive Grammar was presented in Langacker (1982), but the fundamental piece of work was published in two companion volumes: one in 1987—an introduction to the main theoretical tenets and tools—and the other in 1991a—a descriptive application mainly based on English data.

Through the years, Langacker continued to refine and further articulate this theory. Cognitive Grammar has been applied to a wider array of linguistic phenomena, for example, locatives, agreement, passive voice, tense and aspect, complementation, grammaticalization, discourse, and to a wider variety of languages and language families such as Cora, German, French, Greek, Samoan, Slavic, Spanish, just to name a few. For his work, Ronald Langacker is considered one of the leading figures in Cognitive Linguistics, a productive and influential theorist within this framework and in linguistic theory in general.

Biography

Ronald Langacker was born in Fond du Lac, Wisconsin, on December 27, 1942. He graduated from the University of Illinois in 1963 in French. He received his doctorate in Linguistics in 1966 from the same university for his work on generative syntactic theory. He moved to the University of California at San Diego in 1966, where he has been teaching ever since. Currently, he is Professor Emeritus and Research Professor. He is a member of the Program in Cognitive Science at UC San Diego (1983–2003). He is also Visiting Professor and Fellow at a number of institutions all over the world such as University of Chicago (1969), University of Illinois (1969), Belgrade University (1971), Summer Institute of Linguistics (1976), Australian National University (1991), University of New Mexico (1995), Graduate School of Linguistics, Amsterdam (1996), Graduate School in Language and Communication, Odense University (1998), Chulalongkorn University, Bangkok (2001), and University of California, Santa Barbara (Edward Sapir Professor, 2001). He served on both the Advisory Board (1989–1997) and the Governing Board (1999–present) of the International Cognitive Linguistics Association and as President of the same association (1997–1999). He is a recipient of several awards such as the National Endowment for the Humanities (senior fellowship, 1973–1974), Guggenheim Fellowship (1978–1979), the Medal of the University of Helsinki (1996), and an honorary doctorate from the University of Lodz, Poland (2003). Langacker was co-editor of the monograph series Cognitive Linguistics Research and a member of numerous editorial and advisory boards.

References

Croft, William, and Alan D. Cruse. Cognitive linguistics. Cambridge: Cambridge University Press, in press.

Cuyckens, Hubert, and Dirk Geeraerts. Handbook of cognitive linguistics. Oxford: Oxford University Press, forthcoming.

Langacker, Ronald. 1969. On pronominalization and the chain of command. Modern studies in English, ed. by Sanford A. Schane and David A. Reibel, 160–89. Englewood Cliffs, NJ: Prentice-Hall.

———. 1968. Language and its structure. New York: Harcourt Brace and World, 2nd edition 1973.

———. 1972. Fundamentals of linguistic analysis. New York: Harcourt Brace Jovanovish.

———. 1976. Non-distinct arguments in Uto-Aztecan. Berkeley and Los Angeles: University of California Press.

———. 1977. Studies in Uto-Aztecan grammar, Vol. 1, An overview of Uto-Aztecan grammar. Dallas: Summer Institute of Linguistics and University of Texas at Arlington.

———. (ed.) 1979. Studies in Uto-Aztecan grammar, Vol. 2, Modern Aztec grammatical sketches. Dallas: Summer Institute of Linguistics and University of Texas at Arlington.

———. (ed.) 1982. Studies in Uto-Aztecan grammar, Vol. 3, Uto-Aztecan grammatical sketches. Dallas: Summer Institute of Linguistics and University of Texas at Arlington.

———. (ed.) 1984. Studies in Uto-Aztecan grammar, Vol. 4, Southern Uto-Aztecan grammatical sketches. Dallas: Summer Institute of Linguistics and University of Texas at Arlington.

———. 1987. Foundations of cognitive grammar, Vol. 1, Theoretical prerequisites. Stanford: Stanford University Press.

———. 1991a. Foundations of cognitive grammar, Vol. 2, Descriptive application. Stanford: Stanford University Press.

———. 1991b. Concept, image and symbol: the cognitive basis of grammar. Berlin and New York: Mouton de Gruyter.

———. 1999. Grammar and conceptualisation. Berlin and New York: Mouton de Gruyter.

———. 2001. 'Discourse in cognitive grammar'. Cognitive Linguistics 12. 143-88.

Lee, David. 2002. Cognitive linguistics. Oxford: Oxford University Press.

Taylor, John. 2003. Cognitive grammar. Oxford: Oxford University Press.

Ungerer, F., and H.-I. Schmid. 1997. An introduction to cognitive linguistics. London: Longman.

IRAIDE IBARRETXE-ANTUNANO

Language Change

One of the most fundamental linguistic universals is that all languages change over time. This fact is quite obvious in the case of languages having a documented history over many centuries, such as English. An English speaker attempting to learn Old English will quickly find the task equivalent to learning a foreign language. For example, in Old English the word *knight* (originally spelled *cniht*) was pronounced [kniçt]; hence, it contains both a foreign cluster (word-initial [kn]) and a foreign sound ([ç]). Old English was also very different in its word structure. The genitive plural form of *knight* (as in *the knights' horses*) was *cnihta* [kniçta], and there were specific dative forms (e.g. plural *cnihtum* [kniçtum] "to the knights") that no longer exist. There has also been semantic change, since in Old English the word *knight* meant "boy, youth, servant".

But even without historical documents, we are surrounded by various types of evidence showing that English has undergone significant change. Since spelling conventions are often maintained in spite of sound changes, written forms can reflect earlier pronunciations. We still write *knight* even though the *k* and the *gh* have not been pronounced for more than 300 years. Another type of evidence for language change is found in the comparison of English with other Germanic languages. Accordingly, the similarity of English *knight* and German *Knecht* [kneçt] can be explained by the fact that both words come from the same source; namely, an earlier West Germanic language that later split into individual dialects that eventually developed into English and German. In this case, German is more conservative than English in preserving the pronunciation of both [kn] and [ç], and the meaning of the word ("farm laborer, servant") is closer to that of Old English. Another type of evidence for language change is found in idioms and compounds, which can sometimes preserve older features of the language; e.g. *the quick and the dead* and *quicksilver,* where *quick* reflects the Old English meaning "living".

The seeds of language change are found everywhere in the form of linguistic variation. An interesting sociopolitical phenomenon is the popular reaction, almost universally negative, to linguistic variants or perceived changes. It is very often the case that reduced pronunciations are denounced as sloppy, new word formations are rejected as slang or substandard, and words being borrowed from another language are treated as a form of contamination. Such judgments have a long tradition, and no doubt when some people first began pronouncing *knight* with [n] at the start instead of the "correct" [kn], some eyebrows were raised; but over time, a linguistic community adopts one variant over another. Usually, this adoption is done at the unconscious level, but sometimes linguistic markers become prestigious or stigmatized in a fairly conscious way; e.g. whether you pronounce the first vowel in "drama" as [æ] or [ɑ], the last consonant in "garage" as [dʒ] or [ʒ], etc. While such variation might seem haphazard or random, both the variation and the resulting language change have highly systematic properties, and the same patterns of language change are found universally as a natural consequence of our common cognitive and physiological endowment.

Language change can have both external and internal motivations. External motivations involve factors independent of the language system itself, such as borrowing. Languages in contact can borrow extensively from each other, as is evident in English, where the effects of borrowing are most obvious in the case of vocabulary, but in fact all subsystems (pronunciation, grammar, meaning, etc.) can undergo change as a result of contact. Language contact is also of interest since it seems to be a catalyst of change. Languages with less contact tend to be conservative (e.g. Icelandic) and retain many archaic features. Languages with extensive contact tend to be innovative (e.g. English), i.e. they develop away from the ancestral form. Another external type of change

involves the impact of writing on pronunciation (e.g. the word *often* pronounced with a [t] is a spelling-induced pronunciation), although this impact is usually much less than people normally assume. However, even in the absence of such external factors, languages undergo change. A primary factor behind such internally motivated change is found in the language acquisition process itself. A grammar is not something that is handed down from one generation to the next, but each generation—or more correctly, each child—must construct a grammar anew. In the process, the child's grammar can differ subtly from the adult target grammar, and through the accumulation of such differences over many generations, languages can change significantly.

Explanations of particular changes are often based on reference to relative preferences relating to linguistic structure. Such an approach works particularly well in the case of change in pronunciation and word structure. For example, the elimination of word variants (*thrive/throve > thrive/thrived; old/elder > old/older*) is attributed to a preference for "one meaning, one form", and the common loss of a final consonant in syllables is motivated by the preference for syllables consisting of only a consonant and a vowel (CV). This type of explanation is strengthened in cases where the preferences find independent support in areas like language acquisition (e.g. CV syllables are the first to be acquired by children), and implicational universals (all languages have CV syllables, and some languages have only CV syllables).

From a historical perspective, it is often possible to identify drifts; that is, the tendency for a particular language or language family to display interrelated or similar changes over a long period of time. Such drifts may be attributed to the assumption that the same preferences are at work, or that a particular conditioning factor is present over an extended period; e.g. it is often claimed that the fixing of the stress accent on the root syllable in early Germanic led to the weakening and loss of unstressed syllables that is evident in all the modern Germanic languages. Accordingly, languages can drift in the same direction, even after they have become separated. For example, "umlaut" (a change in which vowels become more alike, which is the source of alternations such as *foot/feet*) occurred in both the West and North Germanic languages. Although umlaut itself cannot be attributed to Proto-Germanic, it is apparent that favorable conditioning factors were inherited in both the western and northern branches.

The study of language change is the subject of historical or diachronic linguistics. The following is a brief overview of the main subfields.

Phonological Change

Sound changes are typically triggered by a conditioning environment. A very common type of change is "regressive assimilation", where a segment becomes more similar or identical to one that follows it ([kt] > [tt] as in Latin *noctis*, Italian *notte* "night"—the arrow > indicates the direction of change). Changes can also be conditioned by properties of larger structures such as the syllable; e.g. segments are often deleted at the end of a syllable (Spanish *cabo* "end" shows the loss of [t] in syllable-final position; cf. Latin *caput*). However, some changes, such as "deaffrication" ([t^s] > [s] or [tʃ] > [ʃ]), do not necessarily require a particular conditioning environment, but rather they occur in all environments. Other changes, such as "diphthongization" (e.g. [u] > [aw], [i] > [aj]), appear to find their genesis in the phonological system itself. Vowel systems that include seven or more vowels are highly prone to diphthongization, while smaller vowel systems are not. Although most changes can be described in articulatory terms (e.g. manner or place assimilation, where fewer articulatory movements are required), some types of change show the importance of auditory factors (e.g. [x] > [f] as in *laugh*, where the *gh* was originally pronounced [x]).

Sound changes are usually regular in the sense that the particular change applies to all words displaying the necessary phonetic conditions regardless of its part of speech. However, in some cases a particular change can generalize from one word to another over time. Such "diffusion" is evident in the case of stress shifting in English nouns; e.g. stress has been shifted in *rebel* (formerly *rebél*), but not in *mistake*. Nouns such as *address* and *adult* are subject to significant variation (both *addréss* and *áddress*).

Some changes involve interrelated shifts of entire series of segments or sets of natural classes. One of the most famous sound shifts, called Grimm's Law after its nineteenth-century discoverer Jakob Grimm, involves a set of sound changes that occurred in early Proto-Germanic (the assumed common language from which the known Germanic languages developed). In its traditional formulation, Proto-Indo-European voiceless plosives shifted to voiceless fricatives, voiced plosives to voiceless plosives, and voiced aspirated plosives to voiced plosives; e.g. Latin *pater* maintains the original voiceless plosive [p], whereas the English cognate *father* with the voiceless fricative [f] shows the effects of Grimm's Law.

Morphological Change

Languages are often classified according to the type and complexity of the word structure they have.

"Analytic" languages have very simple words, while "synthetic" languages are morphologically complex—have complex words. Thus, English, a fairly analytic language, does not have a future tense marker or case inflections, whereas a synthetic language such as Latin has both. Over time, languages can lose complexity, primarily through sound change; e.g. the case system of Old English was lost in this way. At the same time, new morphological complexity can develop through morphologization or grammaticalization. "Morphologization" refers to the case where an original sound change becomes a morphological regularity; e.g. the source of irregular plurals such as *foot/feet* can be found in an earlier sound change (umlaut). In one type of "grammaticalization", an original word can become a grammatical item such as a suffix; e.g. the future tense morpheme *-ò* in Italian *finirò* "I will finish" derives from an original auxiliary verb, *habeo*, as in the Latin main verb–auxiliary construction *finire habeo*. Forms undergoing grammaticalization are typically subject to extensive phonological and semantic change.

Leveling and analogy are important kinds of morphological change. "Leveling" refers to the elimination of word variants or "allomorphs" within a paradigm. German has allomorphy in verbal paradigms of the type *ich helfe* "I help," *es hilft* "it helps" (i.e. *helf-* vs *hilf-*). Old English had the same type of allomorphy, but it was leveled out so that Modern English has only *help-* as in *I help, it helps*. "Analogy" is the extension of a regularity from one context to another, especially from members of one paradigm to those of another. For example, in earlier English there were numerous "strong" verbs of the type *drive/drove/driven, sing/sang/sung,* etc. Over time, many of the strong verbs have joined the larger *-ed* verb class through analogy. Thus, *molten,* the original past participle of *melt,* was replaced by *melted* through analogy with *-ed* verbs such as *belt/belted, work/worked,* etc. Sometimes the older forms are maintained in the language with a specialized meaning (e.g. *molten steel*) or they may be completely lost (e.g. *throve,* the original past tense of *thrive*).

"Folk etymology" is an interesting kind of morphological change involving a false or incorrect analysis. The word *bridegroom* traces back to an original compound consisting of the words *brȳd* "bride" and *guma* "man". Over time, the independent word *guma* was lost, and when the original compound was no longer understood, folk etymology occurred, and the word was reanalyzed as "bride" + "groom". Often, the result of folk etymology is semantically incoherent; e.g. the dialectal reanalysis of *asparagus* as *sparrow grass*.

Syntactic Change

Changes in sentence structure are also readily apparent in the history of English. The word order in a sentence such as *ond heo hine in þæt mynster onfeng* (literally "and she him into the monastery received") was entirely acceptable in Old English, but is not in Modern English. Indeed, a variety of evidence supports the reconstruction of the Subject–Object–Verb word order for Proto-Germanic. One of the oldest Germanic sentences, written on the Golden Horn of Gallehus about 1,700 years ago, displays Subject–Object–Verb order: *ek hlewagastiR holtijaR horna tawido* (literally "I Hlewagistir of Holt horn made") "I Hlewagistir of Holt made the horn."

Attempts to explain the shift from Old English S(ubject)–O(bject)–V(erb) to Modern English Subject–Verb–Object are usually grounded in typological studies that established two basic word order types, "dependent–head" (also called "prespecifying" or Object–Verb languages) vs. "head–dependent" ("postspecifying" or Verb–Object languages). Dependent–head languages such as Korean display the following types of orders: objects precede verbs, main verbs precede auxiliaries, relative clauses precede their nouns, and noun phrases precede their adpositions (or, in other words, dependent–head languages have postpositions). Head–dependent languages such as Maori display the opposite word order patterns (objects follow verbs, main verbs follow auxiliaries, etc.). Much syntactic work assumes the primacy of the order of the verb and the object. If the OV order changes to VO, then the head–dependent order can be generalized to the other patterns. From this perspective, it is essential to find reasons for the OV > VO shift. One very old theory that has stood the test of time is based on Wackernagel's Law, which observes that certain particles called "sentence clitics" have a tendency to be placed in clause-second position. If these clitics are verbal elements, then the clause-second position can be reinterpreted as the position of the verb, and (S)VO can result as the default order.

Semantic Change

Change of meaning also displays a systematic aspect, and patterns of change can be clearly identified. Words for "see" often take on a new meaning of "know, understand"; e.g. both *vision* (borrowed from French) and *wise* derive from the same Indo-European root. Words involving a sense of touch often take on a new meaning related to the sense of taste (e.g. *sharp*). At the same time, determining the motivations for semantic change has proved a much more elusive task than is the case for other types of change. Nevertheless, general

categories of semantic change can be determined. In semantic "broadening", the meaning of a word becomes more general or inclusive, as in the word *barn,* which derives from an Old English compound *bere* "barley" + *ærn* "house, store" that was originally used to refer to a place where barley was stored. The opposite, and more common, type of change is called "narrowing". A word such as *deer* displays narrowing, since Old English *dēor* meant "animal". The Modern German cognate *Tier* "animal" retains the older meaning. Social and technological factors can also play a role in semantic change; e.g. *write* originally meant "to scratch", since early writing involved scratching letters onto wood or other materials.

Lexical Change

Change in vocabulary primarily involves the loss or addition of words. In many societies, taboo is an important factor in lexical loss; e.g. the Indo-European word for "bear" was lost in earlier Germanic as the result of taboo. More generally, social or technological changes play an important role in motivating the creation of new words. As the need arises, new words develop either through internal sources (e.g. compounding as in *shareware*) or borrowing (*peyote*, from American Spanish).

References

Campbell, Lyle. 1999 Historical linguistics: an introduction. Cambridge, MA: MIT Press.

Croft, William. 2000 Explaining language change. An evolutionary approach. New York: Longman.

Harris, Alice C., and Lyle Campbell. 1995 Historical syntax in cross-linguistic perspective. Cambridge: Cambridge University Press.

Hock, Hans Henrich.1992 Principles of historical linguistics, 2nd edition. Amsterdam: Mouton de Gruyter.

Hopper, Paul J., and Elizabeth Closs Traugott. 1993 Grammaticalization. Cambridge: Cambridge University Press.

Joseph, Brian D., and Richard D. Janda, (eds.) 2003 The handbook of historical linguistics. Malden, MA: Blackwell.

Labov, William. 1994 Principles of linguistic change, vols. 1 and 2. Oxford and Cambridge, MA: Blackwell.

Lass, Roger. 1997. Historical linguistics and language change. Cambridge: Cambridge University Press.

Lehmann, Winfred. 1992 Historical linguistics, 3rd edition. London–New York: Routledge.

McMahon, April M. S. 1994 Understanding language change. Cambridge: Cambridge University Press.

ROBERT W. MURRAY

Language: Contact—Overview

Broadly interpreted, language contact encompasses such diverse experiences as casual contact with a foreign language during travel, study of a second language as a school subject, and the acquisition of two or more languages in early childhood. Because of the power and prestige of their language, many English speakers never need learn another language and may regard monolingualism as the normal state of affairs. For the majority of the world's population, however, some degree of bilingualism (or multilingualism) is the norm; even small-scale nonliterate societies can be remarkably diverse in their linguistic composition. In the modern era, contact with other languages has increased through such phenomena as modern transportation, the Internet, mass communications, the globalization of commerce, and the growth of English as a world language. The study of language contact intersects with many areas of linguistics, and can be approached from the perspective of the society, the individual, or the language.

Early work in language contact and society looked at the typology of multilingual societies, describing each language found according to its status and function. Often in multilingual societies, different languages are used in the same way as different varieties of the same language in monolingual societies; *domain* analysis attempts to classify these functional differences by looking at the class of situations (or domain) in which each language is used. In stable contact situations, multilingualism often endures for long periods, but when contact accompanies social change, multilingualism is often transitory, as one group shifts to the language of the other. A corollary of *language shift* is *language death*, a disturbingly common phenomenon in the modern world.

Languages in contact may be seen as languages in competition, and their status must be regulated by a language policy, whether overtly formulated or residing covertly in the society's linguistic culture (Schiffman

1996). Language status planning must balance the symbolic value of a single indigenous national language against the competing interests of linguistic minorities, especially of sizeable groups that speak languages with important literary traditions. Ex-colonial countries may also have to take into account the prestige of an ex-colonial language and its value of an ex-colonial language as a link to the outside world as well as the cost of developing an indigenous language that has not been used in administrative, educational, technical, and other "higher" domains. Complementing language status policy is educational language policy, whose goal for speakers of a minority language may be to assimilate them to a dominant language or to help them maintain their language; for speakers of the dominant group, policy may aim to provide some degree of competence in another language, often a minority language of the country or an international language of wider communication, such as English.

Turning the focus to language contact and the bi- or multilingual individual immediately raises the issue of the definition of *bilingual.* How well does one have to know a second language before being considered bilingual? Certainly, the person who can use two languages with equal ease in all situations is a rarity, perhaps even an aberration. When one's time is divided between two (or more) languages, differences in competence are almost certain to arise, depending on the exposure to each. More importantly, because languages in a bi- or multilingual society are generally differentiated by domain, the normal expectation is for the individual's competence to be spread unevenly across the society's languages.

There are many paths to individual bilingualism: the languages may be learned sequentially or more or less simultaneously, in a variety of differing contexts, and with varying degrees of community support. There is a long tradition of interest in the organization of language in the bilingual brain, bilingual language acquisition, and cognitive differences between bilinguals and monolinguals. The last issue is particularly acute in relation to educational policy in bilingual societies, where the outcomes of various types of bilingual programes may be under scrutiny.

Bilinguals may alternate between languages, or *code-switch* for a wide variety of reasons: for purely linguistic reasons, such as the "triggering" effect of a loanword; for situational reasons, such as to mark transitions from one domain to another; for discourse reasons, such as to emphasize a point made, to clarify, or to mark a change in topic; and for social reasons, such as to signal a particular identity, to express solidarity or distance. Much work has been done on the structural aspects of code-switching, for example,

trying to establish constraints on where switches may occur.

In addition to purposeful code-switching, bilinguals may incorporate features of one of their languages while speaking the other. This phenomenon, known in various contexts as *borrowing, transfer,* or *interference,* may be due to incomplete mastery of one of the languages and in this way it is similar to certain interlanguage phenomena seen in second-language acquisition. However, it also takes place when speakers are fluent in both languages. Transfer may affect either language and may involve any linguistic feature: pronunciation, word structure, grammar, vocabulary, or semantics. The nature and extent of transfer are partly determined by the sociolinguistic characteristics of the contact situation (Thomason and Kaufman 1988). When transfer features become permanent fixtures of the host language, we speak of contact-induced change. Most languages have borrowed vocabulary from other languages, but the transfer of other features is now recognized to be more common than it was traditionally thought to be. In historical linguistics, a language that has influenced another during the course of language shift is known as a *substratum, superstratum,* or *adstratum* depending upon whether its speakers were socially inferior, superior, or equal to the speakers of the language to which they were shifting. If languages of different families in an entire region come to resemble one another because of contact-induced change, we speak of a *linguistic area* or *Sprachbund.*

Finally, language contact may give birth to new languages in the form of pidgins and creoles. Pidgins arise in restricted contexts for limited communication needs (such as trading in a bazaar) where contact groups have no language in common. They are not the primary language of any group, and in their initial stages they are stripped down to the barest essentials requisite for the communication of content. Over time, they may expand until they become indistinguishable from normal languages. Creoles are primary languages that arise from the nativization of pidgins or develop directly, usually in multilingual contact situations in which children acquiring the emergent language are present at an early stage and play a significant role in shaping it. Although some see creole genesis as a distinct phenomenon (e.g. Bickerton 1981), for others the structure of pidgins and creoles owes much to the structure of the contact languages. Although each context has its individual characteristics, similar transfer phenomena can be seen in stable bilingualism, language shift and death, second-language learning, pidginization, and creolization.

References

Appel, René, and Pieter Muysken. 1987. Language contact and bilingualism. London Baltimore: Edward Arnold.

Baker, Colin. 1996. Foundations of bilingual education and bilingualism, 2nd edition. Clevedon, UK, Philadelphia: Multilingual Matters.

Bickerton, Derek. 1981. Roots of language. Ann Arbor, MI: Karoma Publishers.

Cooper, Robert Leon. 1989. Language planning and social change. Cambridge, New York: Cambridge University Press.

Fabbro, Franco. 1999. The neurolinguistics of bilingualism: an introduction. Hove: Psychology Press.

Hoffman, Charlotte. 1991. An introduction to bilingualism. London, New York: Longman.

Nettle, Daniel, and Suzanne Romaine. 2000. Vanishing voices: the extinction of the world's languages. Oxford, New York: Oxford University Press.

Paulston, Christina Bratt. 1994. Linguistic minorities in multilingual settings. Amsterdam, Philadelphia: John Benjamins.

Romaine, Suzanne. 1995. Bilingualism, 2nd edition. Oxford: Blackwell.

Rot, Sándor. 1991. Language contact. Frankfurt am Main, New York: Peter Lang.

Schiffman, Harold F. 1996. Linguistic culture and language policy. London, New York: Routledge.

Sebba, Mark. 1997. Contact languages: pidgins and creoles. New York: St. Martin's Press.

Thomason, Sara G. 2001. Language contact. Washington: Georgetown University Press.

Thomason, Sarah G., and Terrence Kaufman. 1988. Language contact, creolization and genetic linguistics. Berkeley: University of California Press.

IAN R. SMITH

Language Death

We consider a language to be dead when it is no longer spoken. Languages may die suddenly, with the extinction of their speakers. More commonly, however, languages die gradually, as their speakers shift to another language.

In rare cases, sudden language death results from natural disasters, such as when a volcanic eruption on Sumbawa Island wiped out all the speakers of Tamboran in 1815. More often, sudden language death is the result of human intervention. For example, it is estimated that up to 90% of the indigenous population of North America was eradicated between 1600 and 1800 by diseases carried by the European settlers and their animals. The Tasmanians of Australia and the Beothuks of Newfoundland both became extinct by the nineteenth century as a result of conquest and genocide by the Europeans. In such cases, the languages of these people died with them.

Gradual language death may occur for a number of reasons. A language may fall out of daily use, becoming restricted to ceremonial and formal functions. For example, the Egyptian language Coptic is used only as a liturgical language by the Coptic Christians of Egypt, whose everyday language is Arabic. Typically, however, gradual language death takes the form of language shift, in which speakers of a minority language assimilate to a dominant language. Language shift is promoted or inhibited not by properties of the languages themselves, but rather by a number of social factors, such as the size and economic status of the minority language population and the relative prestige of the two languages. Language shift is characteristic of low-prestige languages, typically spoken by populations that are disadvantaged economically, politically, and demographically. Examples include indigenous linguistic minorities in Europe, such as Basque in France and the Celtic languages in Britain (Welsh and Scots Gaelic), and immigrant communities in North America and Western Europe.

Although multilingualism is a necessary condition for language shift, it is not sufficient, because speakers may choose to remain fully bilingual and to use both languages. Language shift results from situations of unstable multilingualism, in which speakers abandon the minority language for the dominant language, either involuntarily, through political coercion from the state, or voluntarily, to improve their economic prospects. The process of language shift takes two or three generations to complete: whereas the first generation is monolingual in the minority language or speaks the dominant language only as a second language, subsequent generations are bilingual, with the minority language gradually becoming the second language. When language shift is complete, speakers are monolingual in the dominant language.

Language shift has consequences for the structure of the dying language. (Note that many of these consequences are distinguished from normal processes of language contact and change in degree rather than in kind and in their relative rapidity.) First, there is massive and asymmetrical lexical and grammatical interference from the dominant language. For example,

more words are borrowed into the minority language than into the dominant language. Borrowings are integrated less and less into the sound system and grammar as speakers' command of the language declines, reflecting a more general pattern in the dying language, in which all linguistic systems are reduced and variation between speakers increases. Grammatical complexities are reduced, as are the sound systems, especially in sound distinctions that do not exist in the dominant language. For example, in dying varieties of Scots Gaelic, inflected pronouns such as *rium* "with me" are replaced by a form with separate particles, *ri mi.* Finally, the complexity of sentence structure decreases: word order becomes fixed, complex single-word constructions are replaced by multiword constructions, and subordinate and relative clauses are lost.

Language shift also has consequences for patterns of language use. Because minority languages tend not to be institutionally supported, the decline in usage, the increase in variation, and the changes in linguistic structure fail to elicit a puristic reaction from the educational establishment. The domains of use of the language also become progressively restricted, gradually becoming confined to the home or the local community. Because different domains are associated with different linguistic registers, or styles, this restriction leads to a loss of social norms governing the choice of styles and results in monostylistic speech. As language shift proceeds, fewer speakers know the language well and thus offer an inadequate linguistic model for younger speakers. Although these semispeakers may identify themselves as native speakers of the dying language and may in fact have excellent competence in the communicative skills of the language, they are aware that their grammatical and lexical competence is deficient.

There is some debate in the linguistics community about whether and how language shift should be halted or reversed. Although minority language speakers may voluntarily shift to the dominant language for economic reasons, subsequent generations often express regret at the loss of the dying language. Indeed, language death is a tragedy not only because every language is a unique and valuable communicative system, but also because language is so often correlated with sociocultural and ethnic identity. Language death has been halted or even reversed in a number of instances, such as the revival of Hebrew as a spoken language in Israel and the maintenance of Irish Gaelic in Ireland. Although such success stories (relatively speaking) have been aided by a combination of political and educational provisions, the major factor in preventing language death seems to be the degree to which the dying language is emblematic of a social identity that people wish to preserve.

References

Brenzinger, Matthias (ed.) 1992. Language death: factual and theoretical explorations, with special reference to East Africa. Berlin and New York: Mouton de Gruyter.

Craig, Colette Grinevald.1997. Language contact and language degeneration. The handbook of sociolinguistics, ed. by Florian Coulmas. Oxford: Blackwell.

Crystal, David. 2000. Language death. Cambridge: Cambridge University Press.

Dorian, Nancy C. 1981. Language death: the life cycle of a Scottish Gaelic dialect. Philadelphia: University of Pennsylvania Press.

Dorian, Nancy C. (ed.) 1989. Investigating obsolescence: studies in language contraction and death. Cambridge: Cambridge University Press.

Dressler, Wolfgang. 1988. Language death. Linguistics: the Cambridge survey, Vol. IV: language: the socio-cultural context, ed. by Frederick J. Newmeyer. Cambridge: Cambridge University Press.

Nettle, Daniel, and Suzanne Romaine. 2000. Vanishing voices: the extinction of the world's languages. Oxford and New York: Oxford University Press.

Wurm, Stephen A., Theo, and Baumann. 1996. Atlas of the world's languages in danger of disappearing. Paris: UNESCO Publishing and Canberra: Pacific Linguistics.

JAMES A. WALKER

Language Planning

Language planning is the deliberate, future-oriented, systematic change of language form or use, most visibly undertaken by government, in some communities of speakers. Language planning is directed by, or leads to, the promulgation of a language policy by the government or another authoritative body or person.

A language policy is a body of ideas, laws, regulations, rules, and practices intended to achieve the intended planned language change. Language policy may be realized in very formal (overt) language planning documents and pronouncements (e.g. constitutions, legislation, and policy statements) or in informal

statements of intent (i.e. in the discourse of language, politics, and society), or it may be unstated (covert). Formal policy statements may be symbolic or substantive in their intent. Although the field differentiates between language policy (the plan) and language planning (the plan implementation), the two terms are often used interchangeably in the literature.

Language planning is an outgrowth of the positivistic economic and social science paradigms that dominated the three post-World War II decades. Initially called 'language engineering', the discipline emerged as an approach to developing programs for solving language problems in newly independent developing nations. However, by the 1970s, it was apparent that language problems were not unique to developing nations, but were widely applicable to macro (i.e. state-level) language problems and situations. In the 1990s, policy and planning principles have increasingly been applied to microlanguage situations (e.g. to language problems in companies and organizations). Language planning is a major field within applied linguistics, but it is also closely associated with disciplines such as sociolinguistics and education. Language planning subsumes literacy planning, but language planning goals, including those for literacy planning, are often undertaken on their own—not as part of some broader policy.

The practice of overt (explicit) or covert (implicit) language policy and planning may be one of four types: status planning (about society), corpus planning (about language), language-in-education (acquisition) planning (about learning), and prestige planning (about image). Each of these four types of language planning can be realized from two approaches: a policy approach, with an emphasis on form, or a cultivation approach, with an emphasis on language development and use. These eight language planning perspectives are best understood through the goals that planners set out to achieve. Status and corpus planning have been discussed extensively in the literature, with language-in-education planning often categorized as a part of status planning. Little attention has been paid to prestige planning.

Approaches	Policy Planning (on form)	Cultivation Planning (on function)
Types (overt – covert)	Goals	Goals
Status planning (about society)	Status standardization • Officialization • Nationalization • Proscription	Revival • Restoration • Revitalization • Reversal Maintenance Interlingual communication • International • Intranational Spread
Corpus planning (about language)	Corpus standardization • Graphization • Grammatication • Lexication Auxiliary code standard. • Graphization • Grammatication • Lexication	Lexical modernization Stylistic modernization Renovation • Purification • Reform • Stylistic simplification • Terminological unification Internationalization
Language-in-education (Acquisition) planning (about learning)	Access policy Personnel policy Curriculum policy Methods & materials policy Resourcing policy Community policy Evaluation policy	Reacquisition Maintenance Foreign language / Second language Shift
Prestige planning (about image)	Language promotion • Official government • Institutional • Pressure group • Individual	Intellectualization • Language of science • Language of professions • Language of high culture • Language of diplomacy

Based on Hornberger (1994)

Status Planning Goals

Status planning concerns social goals that are external to the language and that a society must make about the environment in which a language is used. Detailed status planning is a relatively neglected activity. The formal goals of status planning may include officialization, nationalization, and proscription. Functional goals may be the revival, restoration, revitalization, maintenance, and spread of languages or the promotion of interlingual communication (international or intranational).

Corpus Planning Goals

Corpus planning is directed at language and represents an attempt to codify, standardize, or modify and elaborate a language. These goals relate to the linguistic form of the languages, e.g. the standardization of the writing system, orthography, grammar, and vocabulary, or to the cultivation of linguistic functions, i.e. lexical or stylistic modernization, renovation (purification, reform, stylistic simplification, or terminological unification), and internationalization. Language renovation may involve language purification, i.e. the removal of foreign (lexical) influences or the adherence to the classical forms and lexicon of a language. Written forms of a language, by definition, are more standardized and purified than the corresponding oral forms. Although the application of technical linguistic skills is central to meeting corpus planning goals, corpus planning also involves making choices or selecting alternatives that have a social aspect and that must be resolved for such planning to be successful.

Language-in-Education (Acquisition) Planning Goals

Language-in-education planning is user related and aims to develop language education programs and teach a language for various purposes. It may relate to the form of a language-learning program, e.g. in policies concerning access, personnel, curriculum, methods and materials, and resourcing policy and/or evaluation, or to the cultivation of language teaching functions, e.g. reacquisition of a former language, maintenance of an existing one, or foreign or second-language teaching. Although language-in-education planning should co-occur with the other planning types, this often does not happen, and language-in-education planning, through schooling, can become the sole language-change agent.

Prestige Planning Goals

Prestige planning is directed at the image a language needs to develop to promote and intellectualize that language. It is conducted by individuals or groups that have or take the responsibility to create the image of the languages, e.g. in language promotion via government institutions or pressure groups. It may aim at enhancing the functional image and status of the language in key language domains, e.g. in intellectualization, the development of a language of science, a language of professions, or a language of high culture. Prestige planning represents a separate range of activities from status, corpus, and language-in-education planning. The latter are productive activities, insofar as they effect a material change. Prestige planning is receptive and influences how the three other planning activities are acted on and received. Prestige or efficiency of organizational impact levels influences the success of a language plan and the uses to which languages are put.

References

Baldauf Jr., Richard. B., and Allan Luke (eds.) 1990. Language planning in Australasia and the South Pacific. Clevedon: Multilingual Matters.

Clyne, Michael (ed.) 1997. Undoing and redoing corpus planning. The Hague: Mouton de Gruyter.

Cooper, Robert L. 1989. Language planning and social change. Cambridge: Cambridge University Press.

Corbarrubias, Juan, and Joshua A. Fishman (eds.) 1983. Progress in language planning. Berlin: Mouton.

Corson, David. 1999. Language planning in schools. Mahwah, NJ: Lawrence Erlbaum.

Eastman, Carol. 1983. Language planning: an introduction. San Francisco: Chandler and Sharp.

Hornberger, Nancy H. 1994. Literacy and language planning. Language and Education 8. 75–86.

Kaplan, Robert B., and Richard B. Baldauf Jr. 1997. Language planning from practice to theory. Clevedon: Multilingual Matters.

Kaplan, Robert B., and Richard B. Baldauf Jr. 2003. Language and language-in-education planning in the Pacific basin. Dordrecht: Kluwer Academic.

Rubin, Joan, and Björn H. Jernudd (eds.) 1971. Can language be planned? Honolulu: East-West Centre and University of Hawaii Press.

Schiffman, Harold F. 1996. Linguistic culture and language policy. London and New York: Routledge.

Spolsky, Bernard. 2004. Language policy. Cambridge: Cambridge University Press.

Wright, Sue. 2004. Language policy and planning: from nationalism to globalism. New York: Palgrave Macmillan.

RICHARD B. BALDAUF, JR

Language Socialization

Language socialization is the process through which new or novice speakers learn about the language and culture. Language socialization includes both socialization through language and socialization to use language.

Historically, theories of socialization have varied in the degree to which the individual is seen as active in their own socialization. Nineteenth-century theories followed Hobbes' notion of the individual as aggressive, selfish, and asocial by nature and saw socialization as the process of reshaping these natural impulses into prosocial feelings and desires. Freudian theory in the early twentieth century emphasized conflict between human nature (the id) and society (the superego). Functionalism viewed the individual as more passive and more socially directed.

A return to an active role for individuals in constructing social order was found in Mead's theory of symbolic interactionalism, which proposed that individuals and society construct one another through social interaction. This perspective influenced the work of phenomenologists, ethnomethodologists, and social interactionalists. Ethnomethodologists also incorporate close examination of the interactional procedures or methods used to construct a sense of shared context or shared realities, and social interactionalists consider socialization to be an interactional display in which a novice language learner is shown expected ways of thinking, feeling, and acting. Thus, social interactions are considered to be sociocultural environments through which language learners come to internalize and gain performance competence in socioculturally defined contexts.

Language intersects with socialization in three distinct domains: (1) language is used to instruct the learner about what to do, think, feel, etc. (e.g. 'smile'); (2) learners are taught what to say (or not) on various occasions (e.g. 'say "Thank you"'); and (3) learners are instructed in certain features of interaction that vary systematically and are correlated with individual or group variables (e.g. 'good little girls don't shout').

First, language is used to instruct the new or novice speakers about what to do, think, feel, etc. When caregivers and other native speakers use language to and in the presence of new or novice speakers, they are providing information or cues concerning what members are doing. Typically, such instruction takes the form of explicit prompting by the caregiver or an other member of a group. For example, an older person will instruct a younger child in what to say by modeling each utterance for the child to repeat. The prompting routine is marked by characteristic linguistic features. Such routines are usually, but not always, initiated by an imperative verb form meaning 'say' or 'do', followed by the utterance to be repeated. Very often, these utterances have a distinct voice quality and intonational contour, which mark them as components of the prompting activity.

Second, new or novice speakers are taught what to say (or not to say) on various occasions. Language plays a major role in the acquisition of activity/event knowledge with many formal and functional features of discourse carrying sociocultural information, including phonological and morphosyntactic constructions, the lexicon, speech-act types, conversational sequencing, genres, interruptions, overlaps, gaps, and turn length. These structures are socially organized and carry information concerning social order. They are also culturally organized and express local conceptions and theories about the world. As new or novice speakers acquire tacit knowledge and competence in the use of these cues, they acquire knowledge and competence in the social organization of activities and events. Thus, through exposure to and participation in language-mediated interactions, new or novice speakers acquire a knowledge of principles of social order and systems of belief (ethnotheories). For example, among certain social groups, some activities and events may have highly predictable discourse structures (e.g. greetings, jokes, ritual insults, teasing, begging, clarification sequences, trick-or-treat routines), whereas others may have more variable discourse organization (e.g. gossip, negotiations, giving advice, explanations, instructions).

Third, new or novice speakers are instructed in certain features of interaction that vary systematically and are correlated with individual or group variables. For example, the acquisition of sociocultural knowledge by children is influenced by their level of cognitive, social, and linguistic development. In other words, the communicative competence expected of the three-year-old child is necessarily different from that expected of a teenager or adult. Similarly, different social situations require different levels of politeness depending on the status of the speakers (e.g. higher, same, or lower than the speaker), the environment (e.g. church, home, school, etc.), or the intent of the speaker (e.g. emphasize group solidarity).

In addition, new or novice speakers are active socializers of others in their environment. Even infants and small children have a hand in socializing other members of the family into such roles as caregiver, parent, and sibling. As such, second children enter a different social environment than do first children; often, first children 'break in' adults as caregivers. As older children, they may further socialize parents into modes of acting and communicating associated with their school and peer-group experiences. This is illustrated by children's recent role as socializers of computer literacy within their respective households.

Although many theories of language acquisition assume that children develop social and cognitive skills through participating in structured cooperative interactions with more mature members of society, it is important to note that all societies do not rely on the very same set of language-socializing procedures. The cross-cultural research of researchers such as Schiefflin and Ochs (1986) has shown that although prompting a child on what to say appears widespread, expanding children's utterances, using leading questions, announcing activities/events for a child, and using a simplified lexicon and grammar to do so are cross-culturally variable. These practices are not, for example, characteristic of adult–child or child–child interactions in traditional Western Samoan households. This evidence of overt instruction is also of interest in the ongoing debate in the literature regarding the degree to which language acquisition is influenced by innate factors (e.g. Universal Grammar).

References

Berko-Gleason, Jean. 1997. The development of language. Needham Heights, MA: Allyn & Bacon.

Coates, Jennifer. 1993. Women, men and language, 2nd edition Harlow, Essex: Longman.

Fletcher, Paul, and Brian MacWhinney (eds.) 1996. The handbook of child language Oxford. UK: Blackwell.

Gallaway, Clare, and Brian Richards (eds.) 1994 Input and interaction in language acquisition. Cambridge, UK: Cambridge University Press.

Schiefflin, Bambi, and Elinor Ochs (eds.) 1986 Language socialization across cultures. New York: Cambridge University Press.

KAREN WOODMAN

Lateralization and Handedness

The term (brain) lateralization derives from the Latin word *latus* (= side of the body) and refers to the fact that the two hemispheres of the human brain are not exactly alike. The concept of lateralization has earlier given rise to the misleading idea that one hemisphere is actually 'dominant' over the other — namely, the hemisphere in which the major capacity for language resides dominating the contralateral hemisphere. This classic concept of '*cerebral dominance'* has been replaced by the concept of '*hemispheric specialization*', with both hemispheres having different general cognitive operational modes and subserving efficient subdivision of complex functions, rather than showing any superiority of one hemisphere over the other.

The earliest known theory of brain lateralization is about 2,400 years old and stated that different hemispheres of the brain control the two functions of sensation and cognition. Although this early theory involved speculations about the two hemispheres being functionally different, up to the 1860s the majority of the scientific community maintained that the hemispheres were structurally and functionally identical. This viewpoint changed dramatically, when the clinical discoveries of Pierre Paul Broca (in 1865) and Carl Wernicke (in 1875) regarding the special role of the left hemisphere in language were made public. Broca took a firm position about the left hemispheric dominance for articulate speech, the comprehension of spoken language, reading and writing, but was careful to add that language was not the exclusive function of the left hemisphere. He further believed the right hemisphere to mediate speech in left-handed people.

It was not before the twentieth century that scientists began to wonder whether Broca might have been wrong about right-hemisphere language dominance in left-handed people. Current evidence from a number of converging sources, notably the high incidence of the language disturbance known as aphasia after left- but not right-hemisphere damage, indicates that the left hemisphere is dominant for comprehension and articulation of language in close to 99 % of adult right-handed people. Sixty-three percent of left-handed and ambidextrous people also have left-hemisphere language, 24 % have bilateral language abilities, and only 13 % of left-handed and ambidextrous people have predominantly right-hemisphere language.

Recent studies showed that some lateralization of function is already present in newborns and is thus established long before puberty. Moreover, certain anatomical asymmetries in the temporal lobe of the brain are already present in the fetus. This is in contrast to early theories, which favored the idea that the two hemispheres were equipotential at birth and for the first two years of life and that a left hemisphere superiority for linguistic functions develops with age. There is also no extensive agreement as to why functions need to be lateralized. A possible interpretation of the cause of lateralization is that it reflects the evolution of systems, which have diversified because of their functional incompatibility. One major example could be provided by linguistic and spatial skills, which are kept segregated in the human hemispheres, perhaps because optimal performance in one of them is detrimental to optimal performance of the other. Another suggestion is that the dominance of the left hemisphere over the right hand and skilled movement preceded its dominance over language.

Because parts of the unsolved problem of functional brain lateralization may stem from the nature of the method used, it is worth considering the primary methods used to study hemispheric specialization. A common approach to study functional brain lateralization is the investigation of patients with unilateral hemispheric lesions. However, the most compelling evidence of functional brain lateralization has been obtained from the results of studies on patients who have either undergone hemispherectomy (removal of the cerebral cortex of one hemisphere) or commisurotomy, a transection of the corpus callosum (the bundle of nerve fibers connecting the two hemispheres of the brain). Commisurotomy permits a direct test of each hemisphere almost working in isolation from the other in these so-called split-brain patients. Studies with these patients revealed a clear dominance of the left hemisphere for language, particularly the syntactic aspects of language and language output. However, the right hemisphere appears capable of understanding a limited amount of words and rudimentary phrases but is only able to produce speech in an extremely limited context. It is not able to recognize when one word is superordinate to another or judging antonyms and it cannot use word order to assist in deciphering meaning. However, it can indicate when a sentence ends with a semantically inappropriate word and it can make grammatical judgments. A further approach to study functional lateralization is represented by studies in normal subjects, using the simple technique of restricting sensory input initially to one or the other hemisphere. This has been done with visual (tachistoscopic presentation) and auditory (dichotic listening) linguistic material. Language lateralization has also been examined through intracarotid infusion of anesthetics (e.g. sodium amytal), which is a technique to allow one to inactivate reversibly and for a brief period one or the other hemisphere (WADA-test). A variety of neurophysiological methods that measure cortical electric or magnetic field activity (electroencephalography, EEG; magnetencephalography, MEG) or cerebral blood flow and metabolic rate (positron emission tomography, PET; functional magnetic resonance imaging, fMRI) were also applied to study hemispheric specialization. Despite the problems that each of these methods have, the best strategy for studying the functioning of the cerebral hemispheres is to use combinations of multiple methods. If the same hemisphere difference is found with multiple methods, it is unlikely to reflect methodological differences.

Efforts to characterize hemispheric differences in terms of oversimplified dichotomies such as the left hemisphere to be the analytic, verbal part of the brain and the right the perceptual, creative, or emotional part have failed. An interesting concept that has emerged from various studies is that functional lateralization is not rigidly contingent upon the stimulus itself but rather upon the way the stimulus is processed. Both hemispheres play active roles during language processing, with each side contributing in a complementary, not exclusive fashion. The left hemisphere, for instance, shows a slight advantage on recognizing detail, whereas the right hemisphere concentrates on the broad background picture. This would explain why left-hemisphere language areas are so good at a precise representation of words (phonology) and word sequences (morpho-syntax), while the right brain seems to supply a wider sense of context, prosody, pragmatics, and meaning. Moreover, the cortical areas specialized for language are not solely concerned with words. Studies of congenitally deaf individuals have shown that the brain areas devoted to sign language are the same as those that organize spoken and heard communication. Such regions are therefore specialized for symbolic representation and communication rather than for spoken language as such. Currently, the relationship between lateralization and language represents a continuing research challenge in neurolinguistics, but a great deal of research focuses on the brain's integrating rather than lateralized abilities.

References

Bishop, Dorothy V. M. 1990 *Handedness and developmental disorder*. London: Mac Keith Press.

Corballis, Michael C., and Stephen E. G. Lea, (eds)*, 1999. *The descent of mind: psychological perspectives on hominid evolution*. New York: Oxford University Press.

Davidson, Richard, J., and Kenneth Hugdahl (eds), 1995. *Brain asymmetry*, Cambridge, MA: MIT Press.

Geschwind, Norman, and Albert M. Galaburda. 1987. *Cerebral lateralization: biological mechanisms associations and pathology*. Cambridge, MA: MIT Press.

Hellige, Joseph B. 1993. *Hemispheric asymmetry: what's right and what's left*. Cambridge, MA: Harvard University Press.

Hiscock, Merrill, Eran Zaidel, Marcel Kinsbourne, Francisco Aboitiz, Andres Ide, and Joseph B. Hellige. 1998. Lateralization of language and communication. *Handbook of neurolinguistics*, ed by Brigitte Stemmer and Harry A. Whitaker. San Diego: Academic Press.

Iaccino, James F., 1993. *Left brain–right brain differences: inquiries, evidence, and new approaches*. Hillsdale: Lawrence Erlbaum Associates.

Kertesz, Andrew (ed) 1994. *Localization and neuroimaging in neuropsychology.* San Diego: Academic Press.

Lenneberg, Eric H. 1967. *Biological foundations of language*. New York: Wiley,

Molfese, Dennis L., and Sidney J. Segalowitz, (eds) 1988. *Brain lateralization in children: developmental implications*, New York: Guilford Press.

Springer, Sally P., and Georg Deutsch. 1981. *Left brain, right brain.* Oxford: W.H. Freemann, , New York: W. H. Freeman and Company, 5th edition, 1998.

Trevarthen, Colwyn B., (ed) 1990. *Brain circuits and functions of the mind: essays in honor of Roger W. Sperry*. New York: Cambridge University Press.

SABINE WEISS

Latin

The Latin language (*Lingua latina*) is a member of the Italic subfamily of the Indo-European family of languages; it was brought into the Italian Peninsula around the tenth century BCE by a wave of immigrants from the North. As for its ties to non-Italic Indo-European languages, it is closely related to Sanskrit and Greek and to the Germanic and Celtic subfamilies. In Italy, Latin was initially the dialect of the minor district of Latium on the left bank of the lower Tiber River, and then it spread in the entire Italian Peninsula, acquiring a number of features from the indigenous languages it replaced. Latin is supposed to have originally been a member of a group of closely related dialects known as Latin-Faliscan, and its success was the consequence of the successful Roman Empire. Its name comes from *Latini*, a group of related tribes who settled in the territory of Latium.

Early Latin inscriptions exist from around the seventh century BCE. The oldest example of Latin consists of a five-word inscription on a fibula dating from the sixth century BCE. The oldest texts clearly in Roman Latin date mostly from the third century BCE, and among them there are rich texts reflecting an earlier, pastoral society. The affluence of fairly continuous and linguistically consistent material that the emerging Latin literature offers has been indispensable in the reconstruction of the ancestor language, Proto-Indo-European.

Latin underwent changes due to several foreign influences, the most important being those of the Celtic dialects in Northern Italy, the Etruscan language in central Italy, and Greek in Southern Italy, the last one being probably the most important, at least for the literary Latin language. The vocabulary from Late and Medieval Latin has been influenced by the Germanic languages in the army terminology, whereas the Arabic influence was most prolific in the case of the scientific language — e.g. *algebra*.

The history of the Latin language may be divided into four periods, similar to the periods of Latin literature. The Early Period (240–70 BCE) corresponds to the time when Latin became the dominant language in Italy. The Golden Age, or The Classical Period (70 BCE–14 CE), marks the age when the Latin language developed into an artistic medium of expression and attained its greatest richness and flexibility, when it became standardized in grammar and vocabulary. During this period, when it also became the language of the Mediterranean area, three types of Latin could be distinguished: Classical written Latin, Classical Oratorical Latin, and Colloquial, or Vernacular, Latin. The Silver Age (14–130 CE) is the period when Latin became the lingua franca of the western part of Europe. During all this time, the language follows a continuous process of development, the most noticeable changes being in the area of vocabulary and syntax. Finally, the Late Period (200–600 CE) is the time when invading barbarian tribes brought many foreign forms and idioms into the Latin language, to the extent that this new form of Latin was called the *Lingua Romana* and was distinguished from the *Lingua Latina,* the classical language of the educated people.

In the Medieval Period, Latin served as the international written medium of communication, as well as the language of science, philosophy, and theology. Approximately two thirds of all medieval European

literature was written in Latin, and the higher society used it in everyday life. Its vocabulary absorbed many words from the local languages to meet the changed intellectual and social conditions. The Medieval (or Low) Latin borrowed new words from various sources, so that new, distinct meanings emerged. In the fifteenth and sixteenth centuries, the New (or Modern) Latin emerged. The writers of the Renaissance produced a new and brilliant Latin literature that was a very close replica of that of the Latin classical writers. Latin was the diplomatic language of communication until the eighteenth century, the teaching language in the European universities until the late nineteenth century, and the language of the Roman Catholic liturgy until the twentieth century. Today, Latin is officially spoken only in Vatican. However, together with Greek, it is the main source for creating scientific terms all over the world.

The Latin (or Roman) alphabet was created in the seventh century BCE on the basis of the Etruscan alphabet. The original Latin alphabet consisted of 21 letters—A B C D E F Z H I K L M N O P Q R S T V X—, but later *Z* was replaced by *G*. The letters *Y* and *Z* were borrowed from the Greek alphabet after the conquest of Greece in the first century BCE. During the Middle Ages, *J, U,* and *W* were added. The final alphabet consisted of 26 letters, as follows: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z. It must be mentioned that in antiquity, the small letter (minuscule) did not exist, although there were several types of the capital scripts, such as the *elegant book capitals*, with a rounded shape, the *lapidary capitals*, used on monuments, and the *rustic capitals*, which were less elaborate, but easier to write. The minuscules developed later through transformations of the ancient capital letters.

Classical Latin had a typically Indo-European sound system, with five long vowels, five short vowels, and plain consonants. Its major characteristic feature is the large number of diphthongs, which later disappeared from the language.

Early Latin had intonational stress on the first syllable of a word, as opposed to the Latin of the Republican and Imperial periods, when the accent fell on either the last or second to the last syllables of the word.

Latin word structure is the result of gradual simplification of the Indo-European original structure. Latin has five inflected (or variable) parts of speech: the noun, the adjective, the pronoun, the numeral, and the verb; and four uninflected (or invariable) parts of speech: the adverb, the preposition, the conjunction, and the particle.

The nouns are divided into five classes, called declensions, according to their endings. A noun can belong to only one of these declensions. The first declension consisted mostly of feminine nouns. The second declension included masculine and inanimate neuter nouns. The third declension consisted of nouns of all genders, but with similar endings. The fourth declension contained masculine and neuter nouns (with one exception), whereas the fifth declension had only a few words, all feminine. Latin nouns were inflected for gender, number, and case. Latin had six cases: the nominative (or subject case: *frater* 'brother'), the genitive (or possessive case: *fratris* 'of the brother'), the dative (or the giving case: *fratri* 'to the brother'), the accusative (or the direct object case: *fratrem* 'the brother'), the vocative case (*frater* 'brother!'), and the ablative case (or the agent case: *fratre* 'by the brother').

All Latin adjectives are in form really nouns; that is, they have approximately the same endings as nouns. They fall into two classes: the first class contains the adjectives based on the first and second nominal declensions, and the second class contains the adjectives based on the third declension. Apart from the adjectives proper, there is the class of pronominal adjectives, which follow the pronominal type of declension. The adjectives have three grades of comparison: positive (the basic form of adjective: *beatus* 'happy'), comparative (*beatior* 'happier'), and superlative (*beatissimus* 'happiest'), the last two being formed by adding specific suffixes to the stem. Adjectives agree with the noun they modify in the categories of gender, number, and case.

Pronouns in Latin are used to substitute nouns, adjectives, and numerals. The pronouns are of the following main types: the personal pronouns (e.g. *ego* 'I', *tu* 'you'), the reflexive (e.g. *se* 'self'), the possessive (e.g. *meus* 'my'), the indefinite (e.g. *aliquis* 'someone'), the relative (e.g. *qui* 'who'), the interrogative (e.g. *quis* 'who?'), and the demonstrative pronouns (e.g. *hic* 'this', *ille* 'that').

The numerals are of Indo-European origin. The cardinal numeral *unus* 'one' was declined like a first- or second-declension adjective, because it was acting in Latin as a pronominal adjective, with distinct feminine and neuter forms. Apart from 20, the numerals do not inflect after *tres* 'three', except when they are turned into regular adjectives. The cardinal numerals (e.g. *unus* 'one') are usually used with nouns in the nominative case. All ordinal numerals (e.g. *primus* 'the first') are declined like regular adjectives of the first or second declension.

Most of the adverbs are basically adjectival forms that derive from the ablative singular forms of the respective adjectives. They are used to determine a verbal action or a quality of an adjective. There are two types of adverbs: primitive and derivative. The

first type consists of proper adverbs referring to circumstances of place, time, and manner (e.g. *jam* 'already'), whereas the second type consists of adverbs that are derived from nouns or adjectives, by means of special suffixes (e.g. *longe* 'far'). Adverbs do not grammatically agree with the verb they match; i.e. they have one fixed form only.

The verbs have an entirely different set of endings from the nouns. They are divided into four classes, called conjugations. The verb conjugations express the following distinctions:

(1) Singular vs. plural: *amo* 'I love' vs. *amamus* 'we love'.

(2) First vs. second vs. third person, both in singular and plural: *amo* 'I love' vs. *amas* 'you love' vs. *amat* 's/he loves'; *amamus* 'we love' vs. *amatis* 'you love' vs. *amant* 'they love'

(3) Imperfective tense (present, imperfect, future) vs. perfective tense (perfect, pluperfect, and future-perfect): *amo* 'I love', *amabam* 'I was loving', *amabo* 'I shall love' vs. *amavi* 'I loved', *amaveram* 'I had loved', *amavero* 'I shall have loved'.

(4) Mood, which includes indicative, subjunctive, imperative, infinitive, gerund, gerundive, and participle: *amo, amem, ama, amare, amandum, amandus, amans*, respectively.

(5) Active vs. passive voice: *amo* 'I love' vs. *amor* 'I am loved'.

So-called 'inchoative' verbs denote the beginning of an action, state, or occurrence. There are three types of inchoative verbs—: primitive (e.g. *nosco* 'to get to know'), verbal (e.g. *pallesco* 'to become pale'), and nominal (e.g. *nigresco* 'to become black')— according to the way in which they are formed: inchoative verbs proper, inchoative verbs derived from other verbs, and inchoative verbs derived from nouns.

'Impersonal' verbs denote the action of an unspecified agent and are used with no expressed subject (e.g. *pluit* 'it rains'). They do not have forms for the imperative mood, and the compound tenses are formed with the participle neuter form.

Participles are adjectives with regular noun/adjective endings added to a verb stem. It must be mentioned that any Latin adjective or participle can be translated as an adjective, but if there is no convenient noun to attach itself to, it is translated as a noun. In Latin, there are three types of participles: the present participle (e.g. *amans*), the past passive participle (e.g. *amatum*), and the future active participle (e.g. *amaturus*).

The gerundive (e.g. *amandus*) is a specialized verbal adjective implying 'oughtness'; it was used either as an attribute with the noun in indirect cases or simply as an adjective of obligation, whereas the gerund is

a verbal noun in the neuter singular form, or more precisely an adjective in the neuter singular serving as an abstract noun.

Classical Latin had a free word order, in the sense that all combinations are possible. However, subject–object–verb seems to be the preferred order in Classical Latin.

In parallel with Classical Latin, Latin developed a spoken vernacular, which was spread by the Roman army throughout the empire. This *sermo plebeius* or *Lingua Romana*, i.e. Vulgar Latin, later developed into the Romance languages. However, it is quite difficult to study this variety, because almost all written materials in Latin used the Classical Latin forms, and thus there is little documentation of *Lingua romana*. One of the most useful works about Vulgar Latin is the 'Appendix of Probus', a list of correct and incorrect word forms from around the third century. One may say that Ancient Spoken Latin is mainly characterized by freedom of syntax, by the presence of numerous interjections, and by the frequent use of Greek words.

On the other hand, in the development of Latin, other characteristics of Vulgar Latin emerged. Thus, vowels lost the distinction between long and short: most of them became short in all Popular Latin dialects. Many diphthongs disappeared, as well as a large number of final vocals and consonants — sometimes all the final syllables — and in general the pronunciation experienced simplification.

The Latin noun declension was simplified over time, and the number of cases decreased. However, a new part of speech appeared: the article emerged in Popular Latin, the indefinite article from the cardinal numeral *unus*, and the definite one from the demonstrative pronouns *ille* and *iste*. The place of the definite article was before a noun in the western dialects, but it followed the nouns in Dacia. The indefinite article invariably preceded the noun. Subject–verb–object became the established word order.

During the long-lasting and wide use of Latin for scholarly and literary purposes, its influence on all the European languages has been enormous. Not only did the Romance languages assimilate the Latin vocabulary, but also all other languages that, directly or indirectly, came into contact with Latin. The legal, scientific, and technological jargons have thousands of words of Latin origin. For instance, English, a Germanic language, inherited words from Latin both directly and through other languages (especially French), so that the English vocabulary consists of almost 66% Latin words. English uses words of Latin origin, such as *cheese, cup, kitchen*, and *street* in everyday life, not to mention the technological terms, perhaps the most common one being *computer.*

References

Allen, W.S. 1970 Vox latina, Cambridge: Cambridge University Press.

Harris, M., and Vincent, N. (eds.), 1988 The romance languages. London: Croom Helm.

Palmer, L.R. 1954. The Latin language, London: Faber and Faber.

Pinkster, H. 1990. Latin syntax and semantics. London: Routledge.

Sharpley, G.D.A.1999. Essential Latin, London: Routledge.

Wright, R.1982. Late Latin and early romance. Liverpool: Cairns.

RADU DANILIUC

Lehmann, Winfred Philipp

Winfred P. Lehmann has defined a role for the comparative-historical study of Indo-European languages as part of the general study of languages and language history.

With the coming of World War II, Lehmann, like many of his contemporaries, put his linguistic skills to work in the service of his country, participating in code-breaking efforts and language instruction. His work as officer in charge of the Japanese language school (1942–1946) resulted in the publication of an early practical grammar of Japanese.

Following his first academic position (Assistant Professor, Department of German, Washington University, St. Louis, 1946–1949), his move to the University of Texas at Austin (1949–), and a Fulbright Research Fellowship to Norway (1950–1951), a sudden need took the Lehmanns from a Fulbright in Egypt to Ankara and the pressing task of creating materials to teach English to Turkish speakers as Director of the Georgetown English Language Program there (1955–1956). Like his work with Japanese, this task involved the production of materials for a language very different from English. It also underlined the critical need to strengthen language teaching in America.

His move to the University of Texas at Austin soon found him building the German Department, as its Chair (1953–1964), into a department ranked first in the nation, and creating a Linguistics Department (Chair 1964–72) ranked third. He then created a new center to teach a wide range of non-European languages and brought numerous colleagues to Texas. To develop new language curricula, Lehmann took advantage of the National Defense Education Act to create programs at the University of Texas (Austin) for the study of South Asian, African, and Middle Eastern languages.

In the 1960s, the field of computational linguistics began to develop. By the mid-1960s, Lehmann was directing the Texas effort to create a German–English translation system, funded for many years by the US Air Force, and later by Siemens Corporation, which marketed a system capable of doing useful first draft translations of technical material such as that produced in quantity by the Common Market. Besides its Machine Translation (METAL) project, the University of Texas' Linguistics Research Center carried out related research on other, often ancient, languages and still does (http://www.utexas.edu/cola/depts/lrc/index.html).

The impact of Lehmann's work brought him international recognition. Among other things, he was awarded the Brothers Grimm Prize during a summer term as Professor at the University of Marburg (1974), and, in recognition of his scholarly and diplomatic abilities, he was chosen to head the linguistic delegation to China when linguists were first invited after the long diplomatic break (see Lehmann 1975). Because of his breadth and depth of contributions to the fields of linguistics, modern languages, and computational linguistics, Lehmann is unique among linguists to have served as president of three very diverse scholarly bodies: the Linguistic Society of America, the Modern Language Association, and the Association for Computational Linguistics.

Lehmann's book, *Historical linguistics* (in its several editions), has been widely translated (see Lehmann 1962, 1973, 1967, 1969, 1992). This work, originally published as a textbook, encoded the ever new directions of a rapidly changing field, as did monographs such as *Proto-Indo-European phonology, Proto-Indo-European syntax*, and *Theoretical bases of Indo-European linguistics*.

Before Lehmann's *Proto-Indo-European phonology* (1952), Indo-Europeanists reconstructed the sounds of the protolanguage as isolated sounds. Lehmann's study of Indo-European phonology as a system of

interdependent sounds underscored the need to recognize sounds as part of a larger system.

Lehmann's *Proto-Indo-European syntax* (1974), reacting to Chomsky's early syntactic work and Greenberg's work on implicational universals, suggested that general syntactic structural systems of language might limit the accumulation of structures that often resulted from comparative-historical reconstruction. When linguists were still discovering much about the range of grammatical structures in the world's languages, this also called attention to the importance of older Indo-European language data and reminded Indo-Europeanists that their data were important for more general questions of language.

His recent books, *Theoretical bases of Indo-European linguistics* (1993) and *Pre-Indo-European* (2002), further explore the relation of Indo-European to typologically diverse structures among the world's languages and go on to explore the role of linguistic paleontology in dealing with prehistory. In essence, Lehmann (2002) pushes the active language type to a period before Proto-Indo-European on the basis of structures that one might agree are only relics in Proto-Indo-European, an otherwise nominative–accusative language.

Lehmann's later work is a unique attempt to describe the language structures of prehistoric speakers whom Lehmann calls 'pre-Indo-European' as opposed to 'Proto-Indo-European'. He brings older Indo-European linguistic data here to bear on the problem of what language was like before Proto-Indo-European. Lehmann's concern with the intellectual and academic context of Indo-European is thus reflected, not only in his administrative impact on the academic institutions of which he was a part but also in the changing focus of his publications.

Lehmann's work has remained firmly rooted in the practicalities of language and culture, be they language teaching, translating, or the question of historical stages of human language. He has challenged the status quo on many fronts and impacted the course of national and international curricula.

Biography

Winfred P. Lehmann was born on June 23, 1916, in Surprise, Nebraska. Growing up in post-World War I Wisconsin, his undergraduate education at Northwestern College (Watertown, WI, B.A. 1936) reinforced his earlier focus on Greek, Latin, and German. As a graduate student at the University of Wisconsin (Madison: M.A. 1938; Ph.D. 1941), he pursued the study of older Indo-European languages such as Sanskrit and Old Irish. In his Old Irish class, taught by Professor Miles Dillon of Trinity University (Dublin), he not only learned the language of poets but also met Ruth Preston Miller, who became his wife and collaborator on Old Irish (Lehmann and Lehmann 1975).

His publications, awards, and various listings in *Who's Who's* are summarized online (http://www.utexas.edu/cola/depts/lrc/general/facultyhomes/lehmann.html).

References

Lehmann, Winfred P. 1952. Proto-Indo-European phonology. Austin: University of Texas Press.

———. 1962. Historical linguistics: an introduction. New York: Holt, 2nd edition, 1973. Japanese edition: Rekishi gengo gaku josetsu. Tokyo: Kenkyusha, 1967. German edition: Einführung in die historische Linguistik. Heidelberg: Carl Winter, 1969. Spanish edition: Introduction a la linguistica historica. Madrid: Editorial Gredos, 1969.

———. 1973. A structural principle of language and its implications. Language 49.47–66.

———. 1974. Proto-Indo-European syntax. Austin: University of Texas Press.

Lehmann, Winfred P., and Ruth P. M. Lehmann. 1975. An introduction to Old Irish. New York: Modern Language Association.

———(ed.) 1975. Language and linguistics in the People's Republic of China. Austin: University of Texas Press.

———. 1992. Historical linguistics: an introduction, 3rd edition. London: Routledge.

———. 1993. Theoretical bases of Indo-European linguistics. London: Routledge.

———. 2002. Pre-Indo-European. (= Journal of Indo-European Studies Monograph, 41.) Washington, DC: Institute for the Study of Man.

CAROL F. JUSTUS

Leskien, August

Leskien, together with Karl Brugmann, Berthold Delbrück, and Hermann Osthoff, formed the so-called 'ursprüngliche Gruppe' (original group) of the neogrammarian movement that emerged around 1875.

In 1876, Leskien's award-winning monograph *Die Declination im Slavisch-Litauischen und Germanischen* (The Declension in Slavic-Lithuanian (i.e. Balto-Slavic) and Germanic) was published, which was to become a

fundamental contribution to the neogrammarian approach. Within this treatise, he presented a critical analysis of both August Schleicher's genealogical theory — stating that the evolution and relationship of languages could be captured in a 'family tree' model — and Johannes Schmidt's wave theory, according to which specific changes within a language spread in a similar pattern as the waves produced by a stone dropped into the water. Leskien rejected Schleicher's idea of language being comparable to a natural organism (and thus linguistics being a natural science) and criticized Schmidt's wave theory for considering the evolution of language as separate from its speakers, their history, migrations, etc. Leskien's statement that sound laws (i.e. changes in the sound system of a language) have no exception was soon considered to be the main principle of the neogrammarian movement. Analogical change (i.e. assimilation of patterns) as the second important mechanism in language change is only to be admitted as a valid explanation, after every conceivable other means was exhausted and a positive condition for analogy could be stated.

Other features still considered typical of Leskien and the other neogrammarians (although this characterization is far from being exhaustive) are the preoccupation with historical linguistics, the atomistic description of linguistic details, and the neglect of content analysis in favor of form analysis. A work of considerable influence on his own and on the linguistic conception of the neogrammarians in general was William D. Whitney's book *The life and growth of language* (1875), which Leskien translated into German (1876; Leben und Wachstum der Sprache). Whitney fully endorsed the neogrammarian position with regard to sound law and analogy, and the influence of his thinking is traceable in Leskien, when he insisted that language cannot be considered separate from its speakers. In the following years, Leskien published important works on the Baltic languages. His study *Der Ablaut der Wurzelsilben im Litauischen* (1884; The ablaut in Lithuanian root syllables) is considered to be the first monograph devoted to a special problem of Lithuanian historical phonetics. Equally important is his work *Die Bildung der Nomina im Litauischen* (1891; The formation of Lithuanian nouns), comprising valuable material for the following generations of researchers. His *Litauisches Lesebuch mit Grammatik und Wörterbuch* (1919; Lithuanian reader with grammar and dictionary) was posthumously published. Leskien is considered to be the pioneer and founder of Baltic studies in Germany. As the first chair for Slavic philology in Germany, Leskien was one of the forerunners of systematic research into Slavic linguistics. From the mid-1870s he devoted himself to the study of Old Bulgarian; his preliminary work focused on the oldest monuments of the Old Bulgarian language, especially the codices (Supraslensis, Zographensis, and Marianus). As a result of this work, Leskien published two books on Old Bulgarian (Old Church Slavonic) that are still valid today: *Handbuch der altbulgarischen (altkirchenslavischen) Sprache. Grammatik, Texte, Glossar* (1871; A Handbook of Old Bulgarian (Old Church Slavonic). Grammar, Texts, Glossary) and his *Grammatik der altbulgarischen (altkirchenslavischen) Sprache* (1909; A Grammar of Old Bulgarian (Old Church Slavonic)). He preferred the term Old Bulgarian rather than Old Church Slavonic, proving that Old Church Slavonic was mainly based on Bulgaro-Macedonian dialects. Like August Schleicher before him, Leskien thus rejected Franz von Miklosich's 'Pannonian' theory, which claimed that the oldest Slavic literary language was of pannonian-slovene origin. Leskien's reviews of Miklosich's works on 'Old Slovene' are not to be seen merely as a criticism of Miklosich's linguistic method, nevertheless, Leskien, underestimating the value of the first complete Slavic etymological dictionary, vigorously criticized Mikolsich's *Etymologisches Wörterbuch der slawischen Sprachen* (1886; Slavic etymological dictionary) and refused to review it. In his two books on Old Bulgarian, as well as in the first volume of his famous *Grammatik der serbo-kroatischen Sprache* (1914; Serbo-Croatian grammar), Leskien, following Schleicher, proposed a classification of the Slavic verb according to its present stem. Leskien contributed relatively little to syntax; although his work on syntax is documented as truly significant by his academic lectures held during the last years of his life, his two books on Old Bulgarian contain almost nothing and the manuscript of the second volume of his Serbo-Croatian grammar, which is supposed to have dealt with syntax, got lost under the bombing of Leipzig in World War II. Together with Vatroslav Jagic, Leskien in 1876 founded the journal *Archiv für slavische Philologie.* Besides his major works, Leskien published numerous articles on the phonology and morphology of the Slavic languages, as well as on graphematic questions and Old Bulgarian literature. Although Leskien devoted most of his scholarly activities to the linguistic investigation of languages, he was also deeply interested in all cultural aspects of language. His numerous publications on fairy tales, folk songs, and related literary topics (mostly dealing with Baltic, Slavic, and Balkan languages) were widely accepted as equal in quality to his linguistic works. Undoubtedly, Leskien's research laid the foundations of modern Slavic and Baltic studies and left its imprint on historical and comparative linguistics in general. Both as a scholar and as a university teacher, he influenced not only many European but also American linguists. Among those who studied under Leskien, or at

least attended his lectures, were Jan Baudouin de Courtenay, Aleksandar Belić, Leonard Bloomfield, Erich Berneker, Olaf Broch, Ben'o Conev, Filipp F. Fortunatov, Jooseppi J. Mikkola, Holger Pedersen, Ferdinand de Saussure, Aleksandur Teodorov-Balan, Lucien Tesnière, Nikolai S. Trubetzkoy, Karl Verner, Gustav Weigand, and others. Many of his works, especially his *Handbook* and his *Grammar of Old Bulgarian* as well as his *Serbo-Croatian Grammar* have become classics and are still valid and stimulating.

Biography

August Leskien was Born in Kiel, Germany on July 8, 1840. He completed high school in 1860 and began his university studies of Classical philology at Kiel University under Georg Curtius; in 1862, he followed his teacher to Leipzig, Germany, where he obtained his doctorate in 1864. In 1864–1866, he worked as a teacher of Latin and Greek at the 'Thomasschule' in Leipzig; he continued his academic studies of Comparative Indo-European Linguistics with special emphasis on Baltic and Slavic languages under August Schleicher at Jena, Germany, and received his final academic degree *venia legendi* (habilitation) in 1867. In 1867–1869, he was university lecturer of Comparative Linguistics at Göttingen, Germany, and in 1869, succeeded Schleicher at Jena as *extraordinarius* (associate professor) of Comparative Linguistics and Sanskrit. In 1870, he was appointed *extraordinarius* of Slavic Philology at Leipzig and as *ordinarius* (full professor) at Leipzig University in 1876. In 1875, he was Fellow of the Royal Saxonian Society of Sciences; a member of the Jablonowski Society of Sciences in Leipzig and Foreign Member of the Imperial St. Petersburg Academy of Sciences in 1876; a member of the Bulgarian Learned Society (since 1911 Bulgarian Academy of Sciences) in 1884; Foreign Member of the Imperial Academy of Sciences in Vienna in 1898; Foreign Member of the Prussian Academy of Sciences in 1900; a Member of the Swedish Academy of Sciences in Uppsala in 1914; and Foreign Member of the Academies in Belgrade, Munich, and Zagreb. He received an honorary Doctorate at the University of Oslo; Serbian Order 'Sava' Second Degree, Order of the Saxonian King and Title 'Geheimer Hofrat' (Court Councillor). Leskien died in Leipzig on September 20, 1916.

References

Eckert Rainer. (1981). Zu Leskiens baltischen Forschungen. Zeitschrift für Slawistik 26. 199–204.

Eichler, Ernst. (1981) August Leskiens Wirken für die Slawistik. Zeitschrift für Slawistik 26. 168–91.

Eichler, Ernst, and Gerhart Schröter (1983) Nachträge zur Bibliographie der Veröffentlichungen August Leskiens, Zeitschrift für Slawistik 28. 280–83.

Jagic, Vatroslav. (1917) August Leskien. Almanach der kaiserlichen Akademie der Wissenschaften in Wien 67. 481–99.

Jankowsky, Kurt R. 1972. The neogrammarians. The Hague and Paris: Mouton.

Leskien, August. 1871. Handbuch der altbulgarischen (altkirchenslavischen) Sprache. Grammatik, Texte, Glossar; 2nd revised edition 1886, 10th edition 1990.

———. 1876. Die Declination im Slavisch-Litauischen und Germanischen.

———. 1882. Litauische Volkslieder und Märchen. Aus dem Preussischen und dem Russischen Litauen, gesammelt von A. Leskien und K. Brugmann.

———. 1884. Der Ablaut der Wurzelsilben im Litauischen.

———. 1891. Die Bildung der Nomina im Litauischen.

———. 1909. Grammatik der altbulgarischen (altkirchenslavichen) Sprache.

———. 1914. Grammatik der serbo-kroatischen Sprache. 1. Teil: Lautlehre, Stammbildung, Formenlehre.

———. 1915. Balkanmärchen. Aus Albanien, Bulgarien, Serbien und Kroatien.

———. 1919. Litauisches Lesebuch mit Grammatik und Wörterbuch.

Morpurgo Davies, Anna. 1996. La linguistica dell'Ottocento. Bologna: il Mulino; as Nineteenth-century linguistics. London and New York: Longman, 1998.

Schröter, Gerhart. (1981). Bibliographie der Veröffentlichungen August Leskiens. Zeitschrift für Slawistik 26. 271–78.

Streitberg, Wilhelm. 1919 [1921]. August Leskien. Indogermanisches Jahrbuch 7. 138–43.

Whitney, William D. 1875. The life and growth of language; as Leben und Wachstum der Sprache, transl. by August Leskien, 1876.

ALEXANDER SITZMANN

See also **Baltic Languages; Brugmann, Karl; Historical Linguistics; History of Linguistics: Overview; Old Church Slavonic; Serbo-Croatian and South Slavonic Languages**

Lesser Antillean French Creole

Lesser Antillean French Creole is spoken in the Caribbean on the islands of Guadeloupe, Martinique, Dominica, and Saint Lucia. It has been argued to be related to other Atlantic French Creoles such as Guyanese, Haitian, and Louisianais, but such relationships being controversial, we will not take into account

these other creoles in the present study. Within this four-island cluster, at least two major distinctions differentiate the varieties spoken in Guadeloupe and Martinique from those spoken in Dominica and Saint Lucia. First, Guadeloupean and Martiniquais coexist with the French language, whereas the varieties spoken in the Commonwealth Antilles on the islands of Dominica and Saint Lucia coexist with English. Second, Lesser Antillean emerged in Guadeloupe and Martinique (where a plantation colony was developed) and was subsequently 'exported' to Dominica and Saint Lucia by migrants. Given that this article is only concerned with the study of Lesser Antillean French Creole, only the varieties spoken in Guadeloupe and Martinique will therefore be examined. In 1946, Guadeloupe and Martinique became French 'Départements d'Outre-Mer', a status that they still have today despite separatist attempts during the 1960s and 1970s.

Lesser Antillean French Creole developed in the seventeenth century at the onset of the French colonization of Guadeloupe with African slaves in 1635, the objective being the development of a plantation colony based on the cultivation of sugar cane. The influx of Black slaves lasted up to the 1860s. Geographically speaking, African slaves in the Caribbean came from Kwa-speaking areas stretching from Sierra Leone to the Ivory Coast. They also came from Togo, Benin, Nigeria, Ghana, Gabon, Congo, Zaïre, and Angola (Gadelii 1997:2). Parkvall (1999) argued that Lesser Antillean originated in present-day St.Kitts with two main substrata: Kwa and Western Bantu languages. Singler (1995) takes a similar stand, but suggests that the Bantu substratum was more important than the Kwa one based on demographic data. Lefebvre (1986) has defended that the Gbe cluster (including Ewe and Fon) in present-day Togo and Benin constitute the main substratum in the French colonies. The other languages having contributed to the genesis of Lesser Antillean French Creole are not only varieties of seventeenth-century French but also the languages of the Caribs, the indigenous group living in Guadeloupe prior to the arrival of the French colonists. While some scholars do not fully acknowledge the input of Carib languages to the genesis of Lesser Antillean due to the extermination of the tribes by the end of the seventeenth-century (Cérol 1997), others like Wittman and Fournier (1994) think otherwise. According to these authors, the language of the Karina tribe that the French colonists first encountered in the 1620s was heavily influenced by Tupi and may have given way to a French Carib pidgin (Gadelii 1997:37). They further suggested that Arawakan languages spoken by indigenous Ameridians before the Europeans arrived also influenced French-related creoles in the Caribbean.

At the phonological level, the vocalic inventory is as follows: /i/, /e/, /ɛ/, /a/, /o/, /u/ with a little variation in the acrolectal varieties. Lesser Antillean contains some consonants that do not exist in French: /h/, / N /, /tΣ/, and /dZ/. Palatalization is also quite common as in [ky], [gy], [ty], [dy], and [sy] (Gadelii 1997).

While the influence of French on the lexicon of Lesser Antillean French Creole is obvious, a number of parallelisms between Guadeloupean and African languages, documented by Cérol (1997), are worth noting at the lexical level. They fall into distinct categories:

Nouns: Kikongo *musulongo* 'the name of a bakongo group' > Guadeloupean *mousoulongo* 'savage'. Kikongo *bebele* 'piece of meat' > Guadeloupean *bébélé* 'dish made out of pieces of meat'.

African words abound in the realms of the fauna and flora: Kikongo *kyokyo* meaning 'a bird' > Guadeloupean *kyo* 'a species of bird'. Kikongo *malanga* 'yam' > Guadeloupean *malanga* 'species of yam'.

Cérol (1997) also provides numerous examples of lexical parallelisms in the domains of anatomy, food, habitat, music, and dance.

Ideophones: Kikongo*, kya-kya-kya* 'onomatopea for laughter' > *Guadeloupean kya-kya-kya* 'laughter or to laugh'.

Grammatical morphemes: Kimbudu 'for, to give' > Guadeloupean *ba* 'for, to give'.

Adverbs: Kikongo *phwele* 'crowd, a number of, a lot of ' > Guadeloupean *pwele* 'a lot of'.

Besides direct lexical inheritances from the French lexifier and the African substrates, Lesser Antillean has naturally produced many innovations of its own such as *palaviré* 'a couple of slaps', *pépa* 'cheap shoes', or *sanpalé* 'deaf'. Furthermore, the language is extremely innovative in the domain of derivational morphology, as it borrows French affixes to make up new words: *makrèl* 'indiscreet woman' + *-é* > *makrélé* 'to mind other people's business' or *pwofité* 'to abuse' +* *-asyon* > *pwofitasyon* 'abuse, tyranny' (Cérol 1997).

In the realm of morphosyntax, the definite article *la* or *a* always occurs in a postnominal position and represents one of the areas where Guadeloupean displays a behavior different from Martiniquais. In Guadeloupean, the definite marker is consistently *la* as in *vwati-la* 'the car'. In Martiniquais however, while *la* occurs after a consonant (or a semi-consonant) as in *tab-la* 'the table', its counterpart *a* must appear in a postvocalic position, as in *loto-a* 'the car' (Bernabé 1987). Damoiseau (1999: 33) further documented in Martiniquais other variants that differ according to the environment. For instance, *lan* appears after a consonant cluster as in *madanm lan* 'the woman', and *an* appears after a nasal vowel as in *pwason an* 'the fish'. From a cognitive perspective, it is

worth observing that definite nouns are prime candidates for plural marking, otherwise the marker does not appear, e.g. *Sé tab la* 'the tables' (Damoiseau 1999:34). As for the indefinite article, it is always found in a prenominal position as in *an tab* 'a table'.

In the realm of question formation, Lesser Antillean has recourse to two strategies: it may front an interrogative word as in *es ou ka travay* 'do you work' or use raising intonation '*ou ka travay*?' 'you work?' Negation is expressed by the morpheme *pa,* which always appears in a preverbal position: *Patrik **pa** lavé loto a* 'Patrick did not wash the car' (Damoiseau 1999).

As far as tense, mood, and aspect markers are concerned, just like most of its congeners, Lesser Antillean makes use of preverbal unbound markers which occur in isolation or in combination to yield diverse tense, aspect, and mood interpretations. It is important to note that the Indo-European concepts of present, past, future... are not directly applicable to the study of creole languages but are nevertheless used as useful descriptive tools. The following examples from Damoiseau (1999) stage a sample of markers that are representative of the Martiniquais variety of Lesser Antillean.

> *Iv **ka** travay* 'Yves works'. The present tense may be expressed with the marker *ka.*
> *Iv ø travay anpil* 'Yves has worked a lot.' The present perfect involves no marker.
> *Iv **té ka** travay lè nou rivé* 'Yves was working when we arrived.' The combination *té ka* expresses past progressive while *té* in isolation may express past or pluperfect.
> *Iv té ø travay isi a lè i té jenn* 'Yves worked here when he was young.'
> *Jou tala Iv té ø travay anpil*: i té las 'That day Yves had worked a lot, he was very tired.'

This last sentence shows that depending on the context, the same marker + verb combination '*té + V*' may yield different interpretations. As in a number of creoles, one can observe that adverbials like *jou tala* 'that day' have an anchoring effect on the event and are instrumental in deriving the relevant temporal and aspectual interpretation. Such areas of investigation display the true complexity of creole grammars.

References

Bernabé, Jean. 1983. Fondal-Natal, Grammaire Basilectale Approchée des Créoles Guadeloupéen et Martiniquais: Approche Sociolittéraire, Sociolinguistique et Syntaxique, vols. 1,2,3. Paris: L'Harmattan.

Bernabé, Jean. 1987. Grammaire Créole, Fondas Kréyol-la: Eléments de base des Créoles de la Zone Américano-Caraïbe. Paris: L'Harmattan.

Cérol, Marie-Josée. 1997. Langue et Identité en Guadeloupe: Une Perspective Afrocentrique. Condé: Jasor.

Confiant, R. Césaire. 1993. Une Traversée Paradoxale du Siècle. Paris: Stock.

Damoiseau, Robert. 1999. Eléments de Grammaire Comparée Français-Créole Martiniquais. Petit-Bourg: Ibis Rouge.

Gadelii, Karl Erland. 1997. Lesser Antillean French Creole and universal grammar. Göteborg: Gothenberg Monographs in Linguistics.

Lefebvre, Claire. 1986. Relexification in Creole genesis revisited: the case of Haitian Creole. Substrata versus universals in Creole genesis, ed. by Pieter Muysken and Norval Smith, 279–300. Amsterdam and Philadelphia: John Benjamins.

Parkvall, Mikael. 1999. A note on the peopling of English St Kitts. St Kitts and the Atlantic Creoles: the texts of Samuel Augustus Mathews in perspective, ed. by Philip Baker and Adrienne Bruyn, 63–74. London: University of Westminster Press.

Singler, John. 1995. The demographics of Creole genesis in the Caribbean: a comparison of Martinique and Haiti. The early stages of creolization, ed. by Jacques Arends, 203–232. Amsterdam: John Benjamins.

Wittman, Henri, and Robert Fournier. 1994. Le Créole Haïtien, Langue Kwa Relexifiée: Vérification d'une Hypothèse "P&P" ou Elaboration d'Astuces Computationnelles? Plurilinguismes 8. 115–139.

MARLYSE BAPTISTA

Lévi-Strauss, Claude

One of the most original and prolific thinkers of the twentieth century, the French anthropologist Claude Lévi-Strauss is associated with the development of a school of thought in cultural anthropology known as structuralism. Challenging the dominating empirical, positivistic and functionalist tradition in social sciences, he viewed cultures as systems and elaborated the principles of the structuralist method, which sought to discover a rationalist universal theory of mind by identifying unconscious and abstract structures underlying the diversity of human cultures. He derived its method from structural linguistics, deeply convinced that 'linguistics was the only social science which could truly claim to be a science and which had achieved both the formulation of an empirical method and an understanding of the nature of the data submitted to its analysis,'

and therefore should be applied as the main model for the anthropological study of diverse cultural categories.

His famous autobiographical work *Tristes tropiques* (1955) provides an insight into early influences during his formative years of life that greatly contributed to his later conceptualization of the structuralist doctrine in anthropology. In addition to his probably innate inclination for formal logic, reinforced by academic training in philosophy, he was also fascinated by geology, Freudian psychoanalysis and Marxist ideas, which all showed him that true reality was never that which was the most manifest and that the process of understanding involved reducing one type of reality to another. Later, when he had come to New York as a refugee during World War II, he began to search for a suitable method to analyze the data on social and family life of Amerindian cultures collected during his own expeditions to central Brazil, as well as the accumulated ethnographic data on North American Indians published by the Smithsonian Institution. During this period, he met the Russian linguist, Roman Jakobson, who introduced him to the work of Ferdinand de Saussure and the basic principles of structural linguistics, thus exerting a decisive influence on his thinking.

Trying to extend the formal structural features of language described by structural linguistics to cultural categories, Lévi-Strauss began to view culture like a grammar, a formal code consisting of symbols and their logical relationships, while the job of the anthropologist was to decipher this code and to reveal the logical workings of the human mind. In search for analogies between language and cultures, he concludes that just as structural linguistics shifts from the study of conscious linguistic phenomena, observed at the level of surface structures (speech), to the study of their deep unconscious infrastructure (language), so the anthropological analysis should place priority on the underlying principles (deep structures) that lie beneath observable everyday human behavior and cultural phenomena, such as kinship rules (surface structure), in order to explain these phenomena. Structural linguistics also introduces the concept of the system of phonemic units that are put together according to certain rules. The basis of its analysis are not these units as independent entities, rather the relationships between them, based on functioning binary opposites that constitute the structure of the system. By analogy, Lévi-Strauss suggests, the true nature of social and cultural phenomena is not in these phenomena themselves, but in the relationships between them. No element has any significance in itself, but only in relation to all elements in the system, so that it is the entire system with its axes of contrasts and oppositions that defines each and every element within it. Like phonemes, kinship terms and patterns, marriage rules, similar prescribed attitudes between certain types of relatives, and other elements of culture are elements of meaning; like phonemes, they acquire meaning only if they are integrated into systems.

By extending the defining contrastive features of phonemes to the analysis of the meaning of cultural categories, Lévi-Strauss suggests that the underlying principles of human behavior and culture also take the form of oppositions, such as nature vs. culture, male vs. female, earth vs. sky, hot vs. cold, etc. However, unlike Saussure, who held that meaning was not a preexisting concept but arbitrarily attached to the form and variable across languages, Lévi-Strauss maintains that all human minds are ultimately the same, the underlying organization of all cultures is the same, generated through the universal set of semantic oppositions, the apparent differences being due simply to varying arrangement of these oppositions. Because of the absolute universality of these principles, what can be discovered by analyzing cultural categories and products in one culture is merely a formal transformation of those in completely different cultures. Accordingly, the focus of anthropological research for Lévi-Strauss is the notion of social structure, which represents a model of actual social relationships and serves to explain them. This structure or model is a system of elements, none of which can undergo change without affecting others. The model has the possibility of predicting change. For every given model, there is a system of transformations that form a set, a group of models of the same type, which make it possible to predict how the model will react if one or more of its elements are submitted to change.

In his first major work, he applied his theoretical concepts to one of the central themes of anthropology, the study of kinship. Contrary to the then prevalent view of kinship patterns as explained only by relations of blood and descent, under the influence of the French anthropologist Marcel Mauss and his exchange theory as central to social structure, Lévi-Strauss bases his theory of kinship on relationships of marital alliance and the exchange of women resulting from rules of incest avoidance. Incest taboos and marriage rules determine whom one should marry and whom one should not, implying reciprocity and the need for exchange while generating marriage structures that determine the distribution of women through marriage to different groups. The exchange of spouses, he says, is the universal basis of kinship systems in the same way the exchange of goods is the basis of the economic system and the exchange of words is the basis of the communication system. Elementary kinship structures are those in which there exists a prescribed, positive rule for marriage to someone of a particular kinship category, for example, to a cross cousin (father's sisters' and mother's brothers' children). Complex kinship structures are those that have negative marriage rules, specifying which persons one may not marry.

According to Lévi-Strauss, the elementary structures of kinship of small-scale societies consist of a system of

relationships, such as the mother's brother/sister's son relationship called avunculate, which involves relationships between mother, father, son and mother's brother characterized by consanguinity, affinity, and descent. If there is a positive relationship (+) between husband and wife, there will be a negative relationship (-) between the wife and her brother, and vice versa. If there is a positive relationship between father and son, there will be a negative relationship between that son and his mother's brother. Lévi-Strauss shows how these relationships vary in different societies depending on whether the rule of descent is patrilineal or matrilineal, using the simple formulaic expression A:B::C:D, as in mother's brother (A) : sister's son (B) :: father (C) : son (D). Proceeding in this way, he reduces diverse historical and geographical rules of kinship to only three elementary structures, constructed through only two types of exchange of women, which themselves depend on a single differential characteristic —- harmonic or disharmonic regime considered, while the whole set of prescriptions and prohibitions can be deduced from the relationship between the rule of descent and the rule of residence.

Lévi-Strauss fully illustrates the structuralist ideas in his analysis of the phenomenon of totemism, the practice of associating people and social groupings, such as clans, with animals, plants, or other objects, related to certain beliefs and rituals. Arguing against the evolutionary view on totemism as a form of primitive religion, as well as against the functionalist explanation of the phenomenon in terms of usefulness of animals, Lévi-Strauss defines totemism primarily as a system of classification of social groups based on the analogy with distinctions between species in the natural world. Totemic classifications, thus, reflect a system of oppositions based on natural objects to organize a set of cultural ones. He opposed the widespread view on presumed mental inferiority of natives and significant differences in the thought processes of so-called "primitive" or non-Western peoples and the modern scientific Western culture. He shows that traditional non-Western peoples with the *"Savage Mind"* have an inductive "science of the concrete", making generalizations from primary sensible data, which "is as rigorous as that of modern science". Close to sensory intuition, the "science of the concrete" differs from Western science in that it operates by working directly through perception and imagination, classifying things in terms of their everyday, sensible features, rather than the underlying abstract features and deductive formal principles that the Western scientist uses. In spite of these differences, Lévi-Strauss claims that the basic thought processes in terms of distinctive features and semantic oppositions remain the same in all individuals, "primitive" or "civilized". These oppositions for Lévi-Strauss are "good to think with" and they distinguish humans from animals, culture from nature.

In his most important work of four volumes of *Mythologiques* or *Introduction to the science of mythology,* he developed the line of enquiry to its climax by analyzing 813 myths taken from 200 South and North American Indians and presenting them as the purest manifestation of the *"Savage Mind"*. He stressed that myths do not have a practical function, but are aimed at solving problems as an end in itself. Comparing myth to language, he borrows from structural linguistics the concepts of syntagmatic and paradigmatic relations (paradigmatic relationships constitute a set in which one item can be substituted for another, and syntagmatic relationships are ones of contiguity and linearity) and describes the workings of myths in terms of relations of opposition, symmetry, substitution, and permutation. Myths function as systems of transformations, so that a single myth can never be understood in isolation, only in its relation to other myths. The events narrated in any single myth form a syntagmatic chain, while the personages and events can be studied as members of paradigmatic sets occurring in a corpus of myths. In his first volume of *Mythologiques, The raw and the cooked,* he begins by analyzing a myth of Bororo people and ends up examining a whole collection of myths from South America, showing as their common structure the discovery of fire and cooking, i.e. the transition from nature to culture, symbolized through invariant oppositions such as raw:cooked (the cooking of culture transforming the raw of nature), or fresh:rotten (the natural process of purification or a return to nature). Within this common structure of nature vs. culture, other analyzed myths show a series of transformations, by substituting e.g. the origin of fire with the origin of water, honey or tobacco. In this way, all the Amerindian myths studied by him are to be considered as variant versions of one another, linked together by transformations, just "as the instrumental parts of a musical work, to be studied as one studies a symphony".

The main criticisms of Lévi-Strauss's structuralism have been concerned with its too formal logic based on literal adoption of linguistic methods, with its negation of history and human initiative, as well as with overemphasis on universals of pan-human unconscious structures and on the intellectual aspect of culture on account of its emotive and practical sides. Despite these various critiques, structuralism represented a dominant intellectual framework in the1960s and 1970s, so that the impressive and voluminous work of Lévi-Strauss has influenced not only further developments in anthropology, particularly in the fields of symbolic anthropology (focused on the interpretation of meaning) and cognitive anthropology (related to mental representations of cultural practices), but also a wide range of intellectuals of varied specializations, labeled post-structuralists, such as the philosopher Jacques Derrida, the intellectual historian Michel Foucault, the psychoanalyst Jacques Lacan, and the literary critic Roland Barthes.

Biography

Claude Lévi-Strauss was born in Brussels, Belgium, on November 28 1908. He completed his B.A. in law and philosophy from the University of Paris in 1931. He held a teaching post in a provincial secondary School, 1932–1934. He moved to São Paulo, Brazil, where he taught sociology at the University of São Paulo, and made several ethnographic expeditions to central Brazil, 1935–1939. He returned to France, where he was mobilized by the French army in 1939–1940. He emigrated to New York in 1941, where he was visiting professor at the New School for Social Research, New York, 1942–1945; he was also Cultural Counsellor at the French Embassy to the United States of America, 1946–1947. He presented a doctoral thesis on 'The elementary structures of kinship' at the Sorbonne, Paris in 1948. He was Deputy Director of the Musée de l'Homme, Paris, 1949–1950; Director of Studies, École Pratique des Hautes Études, Paris, 1950–1974; and Professor and Chair in Anthropology, Collège de France, Paris, 1959–1982. He received the Prix Paul Pelliot in 1949, the Huxley Memorial Medalin 1965, and the Erasmus Prize in 1973; he was also a Member of Académie Française in 1973. After retirement in 1982, Lévi-Strauss currently resides and writes in Paris, France.

References

Barbossa de Almeida, Mauro. 1990. "Mathematical metaphors in the work of Lévi-Strauss." Current Anthropology 31.

Boon, James. 1971. From symbolism to structuralism: Lévi-Strauss in a literary tradition. Oxford: Basil Blackwell.

Charbonnier, Georges. 1961. Entretiens avec Claude Lévi-Strauss. Paris: Plon-Julliard; Conversations with Claude Lévi-Strauss, transl. by John and Doreen Weightman, London: Cape, 1969.

Hayes, Eugene, and Tanya Hayes (eds.) 1970. Claude Lévi-Strauss: the anthropologist as hero. Cambridge, MA: MIT Press.

Hénaff, Marcel. 1991. Claude Lévi-Strauss. Paris: Pierre Belfond.

LaPointe, Francoise, and Claire LaPointe. 1977. Claude Lévi-Strauss and his critics: an international bibliography of criticism (1950–1976) followed by a bibliography of the writings of Claude Lévi-Strauss. New York: Garland.

Leach, Edmund. 1970. Claude Lévi-Strauss. New York: Viking.

Lévi-Strauss, Claude. 1949. Les Structures élémentaires de la parenté; Paris: Puf; The elementary structures of kinship, transl. by James H. Bell, John R. von Sturmer and Rodney Needham. Boston: Beacon Press, 1969.

———. 1955. Tristes tropiques, Paris: Librairie Plon, 1955; Tristes tropiques, transl. by John and Doreen Weightman. London: Jonathan Cape Ltd, 1973.

———. 1958. Anthropologie structurale. Paris: Librairie Plon; Structural anthropology, transl. by Claire Jacobson and B.G. Schoepf. New York: Basic Books, Inc., 1963.

———. 1962. La Pensée Sauvage, Paris: Librairie Plon; The Savage Mind, transl. by George Weidenfeld. London: Weidenfeld and Nicolson, 1966.

———. 1962. Le Totémisme aujourd'hui, Paris: Librairie Plon Totemism, transl. by Rodney Needham. London: Merlin 1964.

———. 1964. Le Cru et le cuit, Paris: Librairie Plon; The raw and the cooked: Introduction to a science of Mythology, Vol. 1, transl. by John and Doreen Weightman. New York: Harper and Row, 1969.

———. 1973. Anthropologie structurale deux, Paris: Librairie Plon; Structural anthropology Vol. II, translated by Monique Layton. Chicago: University of Chicago Press, 1976.

———. 1975. La Voie des masques, Paris: Librairie Plon; The way of the masks, transl. by Sylvia Modelski. Seattle: University of Washington Press, 1982.

Mehlman, Jeffrey. 1974. A structural study of autobiography: Proust, Leiris, Sartre, Lévi-Strauss. Ithaca, NY: Cornell University Press.

Pace, David. 1983. Claude Lévi-Strauss: the bearer of ashes. Boston, MA: Routledge.

Rubel, Paula, and Abraham Rosman. 1996. Structuralism and poststructuralism. in Encyclopedia of cultural anthropology, ed. by David Levinson and Melvin Ember. New York: Henry Holt.

Sperber, Dan. 1985. On anthropological knowledge. Cambridge: Cambridge University Press.

ANITA SUJOLDZIC

Lexical Borrowing

Lexical borrowing occurs when a language 'borrows' a word from another language and incorporates the new word into its own lexicon. The term 'borrowing' is in some ways an unfortunate metaphor, since the donor language is never asked whether it wants to lend the word or not, and the 'borrowing' language never repays the word. However, other potential metaphors such as 'giving' or 'stealing' are equally problematic, and the term 'borrowing' is so well established that we will just accept the term.

Lexical borrowing comes about when two languages are in contact with each other, and there are speakers displaying some degree of bilingualism in both borrowing and source languages. Sarah Thomason and Terrence Kaufman (1988) propose a 'borrowing scale' in which lexical borrowing occurs

with only 'casual contact', while at the upper end of the scale 'very strong cultural pressure' leads to 'heavy structural borrowing'. Lexical borrowing often occurs with only minimal contact. The English word, kangaroo, was borrowed into English from the Australian Aboriginal language Guugu Yimithirr when Captain Cook and his crew were making repairs to their ship on the northeast coast of Australia, and none of the crew ever learned to speak Guugu Yimithirr. Most probably, the word was brought into English after one of the crew pointed at an animal and was told by one of the local Aboriginals that the animal concerned was a 'kangaroo'. In fact, the Guugu Yimithirr word *gang-guru* refers specifically to the large black kangaroo.

More intensive cultural contact than occurred between Captain Cook's crew and the Guugu Yimithirr can lead to much more extensive borrowing. The Norman Conquest of England in 1066 led to English adopting literally tens of thousands of words from French, so that now almost half the words found in an average English dictionary will have their origins in French. Sometimes, the same word has been borrowed more than once, at different times and with different meanings: the English word 'chief' came from Old French *chef* meaning 'head' or 'boss'. The English word 'chef' was borrowed much later from the modern French expression *chef de cuisine* or 'head of the kitchen'.

For borrowing to occur, the borrowing language generally has to be gaining something, or it would probably make do with the words that it already has. Borrowing is often thought to occur for either reasons of necessity or reasons of prestige.

Necessity can explain the borrowing of 'kangaroo' from Guugu Yimithirr, 'banana' from Wolof, or 'kayak' from Eskimo. English has borrowed countless terms for flora, fauna, and other things that were not traditionally found in Britain.

Prestige is often involved in situations where one language is thought by its speakers to have more prestige than the other. This motivation explains all of the French words that came into English after the Norman conquest. Up until the end of World War II, French had borrowed relatively few words from English, but since then French has borrowed countless words from (American) English, referring to many aspects of pop culture and technology. Words like *sweat* 'sweatshirt', *pull* 'pullover', *jean* 'jeans' are used to refer to items of clothing popularized by Americans. Sometimes, a culture can feel threatened by the arrival of large numbers of borrowings. The *Académie Française* has fought a valiant but losing battle against all of these Americanisms, often proposing to replace them with more French equivalents. Few of their proposals have ever had any success.

Some words are more 'borrowable' than others. Basic vocabulary (including body parts, natural geographic phenomena, weather terms, small numerals, and the like) are rarely borrowed. Nouns are more borrowable than other parts of speech. Einar Haugen (1950) reports the following percentages of loans from English into American Norwegian: 75.5% nouns, 18.4% verbs, 3.4% adjectives, and only 1.2% adverbs or prepositions with 1.5% of others.

Borrowing languages can be faced with the problem of how to fit the borrowed words into their own linguistic systems. Two basic strategies are found: adaptation and adoption. Sometimes, different speakers of the same language choose different strategies, as when different speakers of English with varying degrees of familiarity with French use different pronunciations for words like 'croissant'.

Adaptation can be either phonological or grammatical. Many languages do not allow consonant clusters or final consonants, but the words they borrow from other languages may have them. Japanese generally resolves this problem by adding vowels between or after the consonants concerned, as it has with words like *besaboru* from English 'baseball'. In some instances, English homophones have created completely different Japanese words as with *seturoku,* a baseball 'strike', and *seturoki*, an industrial 'strike'.

Grammatical adaptation is required when a noun is borrowed from a language without grammatical gender, for example, into a language that has gender. Most nouns borrowed into French are made into masculine nouns, as were all the Americanisms given above. Verbs that are borrowed into French also require adaptation. The infinitive forms of French verbs all have a suffix of one of the forms -*er, -ir,* or -*re.* Virtually all borrowed verbs end up with the -*er* suffix, as, for example, *chatter* 'to chat in an internet chat room'. Sometimes, grammatical adaptation has remarkable consequences. Swahili is a language with a complex set of noun class prefixes. The prefix *ki-* marks one set of singular nouns while the prefix *vi-* marks their plural counterparts. When the English term "keep left" was borrowed as *kiplefiti*, the initial *ki-* was interpreted as the singular noun class prefix *ki-;* hence, traffic roundabouts (plural) are referred to as *viplefiti*.

Whether a language chooses to adapt or adopt often depends on the degree of familiarity its speakers have with the donor language. Taba (a minority language from eastern Indonesia) borrowed the Malay word *jadi* 'thus' replacing initial Malay /ʤ/ with the closest native equivalent /d/. In more recent times, Taba speakers have borrowed many more words (like *baju* 'shirt') with the Malay sounds preserved so that now,

/ʤ/ can actually be considered to be a native Taba phoneme.

References

Haugen, Einar. 1950. The analysis of linguistic borrowing Language 26. 210–31.

Hock, Hans Heinrich. 1986. Principles of historical linguistics. Berlin: Mouton de Gruyter.

Hock, Hans Henrich, and Brian D. Joseph. 1996. Language history, language change and language relationship: an introduction to historical and comparative linguistics. Berlin, New York: Mouton de Gruyter.

McMahon, April M.S. 1994. Understanding language change Cambridge: Cambridge University Press.

Thomason, Sarah Grey, and Terrence Kaufman. 1988. Language contact, creolization, and genetic linguistics. Berkeley and Los Angeles: University of California Press.

Trask, R.L. 1996. Historical linguistics. London: Arnold.

Weinreich, U. 1968. Languages in contact: findings and problems. The Hague: Mouton.

Winter, Werner. 1973. Areal linguistics: some general considerations. Current trends in linguistics, Vol. 11: diachronic, areal and typological linguistics, ed. by Thomas A. Sebeok. The Hague: Mouton.

JOHN BOWDEN

See also: **Language: Contact—Overview**

Lexicalization

Lexicalization is a term that is used by linguists for referring to several things. The most common uses of this term are the following two: (a) When we create a word to express some meaning, we are 'lexicalizing' that meaning by means of the new term (e.g. the word *software).* (b) Another use of the term 'lexicalization' is found in historical linguistics, where it stands for a process in which a sequence of words is reduced to a single word (e.g. *nevertheless).* In both uses of the term, there is a new lexical item that is introduced to express something for which there was no single lexical item before. Thus, it is possible to find cases of lexicalization in which the word that acquires a new meaning was already available in the dictionary, although it was related to a different concept.

Semantic lexicalization has many causes. The most typical one is the extension of the meaning of existing terms by means of specialization and generalization processes, as well as through metaphor and metonymy. In a process of specialization, the word's original meaning is always narrowed down to a smaller set of special referents. An example of specialization is the English word *corn*, which was originally a term used for 'all kinds of grain'. Later, it specialized to the most typical referent in different countries, so that in the end, it acquired the meanings of 'wheat' in England, 'oats' in Scotland, and 'maize' in the United States. Something similar happened to the English word *queen*, which also went through a specialization process. Its original meaning was any 'wife or woman', but now it is restricted to only one type of wife, that of the king. As for generalization, it is the process opposite to specialization. We find generalization in the extension of meaning that took place in the senses of the word *school.* Originally, a school is 'an institution of learning'. However, this meaning has been broadened to that of 'any group of people mentally engaged upon shared activities or sharing views of style or opinions', giving as a result the senses of 'group of artists' or 'group of people sharing opinions'. Another example of generalization is the word *moon.* Originally, it referred to the Earth's satellite, but it is now applied to any kind of satellite.

Metaphor can be very productive for meaning extension. Metaphor (from Greek *metapherein,* 'carry over') is a process based on perceived similarity. We may use a word to refer to a certain notion because of their similarity of meaning, which does not need to be very close. The important point here is that the similarity should be perceived by the user, whatever it is based on. Eventually, the metaphorical sense will become lexicalized. The interpretation of a homonym such as *school* in the sense of 'group of fish' can be related to the senses of *school* as 'group of learners' and may thus be motivated by the relation of similarity that we perceive between a group of learners following a class and a group of fish swimming together and following a leader.

In metonymy, on the other hand, the extension of meaning is based on a relationship of contiguity, i.e. between the whole of something, e.g. *school* as an 'institution of learning', and a part of it, e.g. the lessons. The word *school* can metonymically stand for each of its components, i.e. the building itself, the lessons, the pupils, the staff, the headmaster, etc. The part-for-whole relationship is one of many possibilities in

metonymy, and there are also many others, such as the producer for product (*I'll have a* **Pepsi-Cola**), material for thing (*Take the* **iron**), effect for cause (*He had a* **long** *face*), etc.

All these are cases of lexicalization of new meanings using already existing words. In some other cases, new words are introduced into the language, by means of borrowing (*plateau* from French*, gestalt* from German, etc.) or creation of new terms (*telephone, software,* etc.). These processes also lexicalize meanings. In any case, both the extension of meaning and the incorporation of new words are processes that can also be examined from a diachronic (historical) point of view, because the new items enter the language at a certain moment of its history and take some time to be firmly incorporated into the system.

An interesting case in this respect is what happens when groups of words that function consistently together become only one word in the course of time. In some cases, they are still used as separate words, even though the expression has to be understood as a whole. This is what happens, for instance, with the Spanish expressions *sin embargo,* 'however', or *por lo tanto,* 'therefore', which have little to do with their components: *sin* ('without'), *embargo* ('embarrassment') would give 'without embarrassment', and *por*

('because of'), *lo* ('the') and *tanto* ('so much') would give something like 'because of (the) so much'. It is interesting to notice here that their English counterparts are also lexicalized expressions, but in this case they have become single words: *however* is made up of *how* and *ever*, and *therefore* of *there* and *fore* (*fore* no longer exists as a single word, but it has given us the present-day English word *before*). Other examples are *notwithstanding*, *nevertheless,* and *alright*. We also have lexicalized compounds among nouns, adjectives, and verbs, e.g. *blackbird, wheelchair, colour-blind, to vacuum-clean*, etc.

References

Bauer, Laurie. 1983. English word-formation. Cambridge: Cambridge University Press.
Cruse, Allan.1991. Lexical semantics. Cambridge: Cambridge University Press.
Dirven, René, and Marjoljin Verspoor. 1998. A cognitive exploration of language and linguistics, [Chapter 2: Morphology.] Amsterdam and Philadelphia: John Benjamins.
Leech, George. 1974. Semantics, Harmondsworth: Penguin.
Lyons, John. 1977. Semantics. Cambridge: Cambridge University Press.
Ullmann, Stephen. 1957. The principles of semantics. Oxford: Blackwell.

CARLOS INCHAURRALDE

Lexicon: Overview

A lexicon (or 'mental dictionary') contains all the words that a person knows; it is a model of how words are organized and stored in the brain. The lexicon specifies how a word is spelled, how it is pronounced, what it means, and its part of speech (e.g. noun, verb, adjective, preposition).

The lexicon figures prominently in grammatical theories, although the exact content of a 'lexical entry' (also called a *lexeme*) differs among theories. Structural theories propose that verb lexemes, for instance, include information about the types of object that a verb might take. According to this approach, the entry for *kiss* indicates that it takes a noun phrase (as in *Mary kissed the kitten*). The subject noun phrase is not specified in the lexical entry because its existence is determined by more general rules. In contrast, lexical theories assume that syntactic information is stored with individual lexical entries rather than as more general syntactic rules. Under this approach, lexemes

specify how particular nouns, verbs, and adjectives combine with arguments. For example, the verb *give* takes three arguments (e.g. the arguments for *Rod gave the cat to Jessica* are *Rod, the cat,* and *Jessica*). The verb's lexeme specifies information about semantic roles associated with the verb (e.g. the entry for *give* indicates that this verb takes an agent and a patient). The entry also indicates the mapping between semantic roles and syntax (e.g. the lexeme for *give* indicates that the agent is the subject and the patient is the object when the sentence is in the active voice).

Words often contain multiple morphemes, i.e. meaningful word parts (e.g. *walking* consists of *walk* and *-ing*, *firefly* consists of *fire* and *fly*). A debate has arisen over whether the mental lexicon contains whole words or only morphemes. Some theorists argue that each variant of a word (e.g. *brush, brushed,* and *brushing*) has its own lexeme. However, this approach is problematic for highly agglutinative languages such as

623

Finnish and Turkish because, in these languages, many prefixes and suffixes can be attached to words. Consequently, the words in these languages can have thousands of variants and would require many lexemes to represent all possible variants of each word. Therefore, some theorists have adopted an alternative approach to the one-lexeme-per-variant view. They propose that the lexicon contains solely morphemes and that variants do not have their own lexeme. Instead, words are decomposed into their morphemic parts (during recognition) or composed (during production). Which view is correct appears to depend on the kinds of morphemes that make up the word. Highly regular morphemes (called inflectional morphemes) do not significantly alter the root morpheme (e.g. the morpheme *-s* indicates plurality but otherwise does not alter the meaning of the word). However, less regular morphemes (called derivational morphemes) significantly alter the root morpheme (e.g. the morpheme *-er* changes *learn* from a verb to a noun). Multimorphemic words with derivational morphemes appear to have their own lexemes, whereas words with inflectional morphemes do not appear to be stored as whole units.

The representation of compounds in the lexicon depends on whether the meaning of the compound is 'transparent' or 'opaque'. Compounds such as *buttonhole* and *snowman* are said to be semantically transparent because their meaning can be derived from the individual morphemes. Other compounds, such as *buttercup* and *hippocampus,* are semantically opaque because their meaning is not derivable from their parts. Semantically opaque compounds (e.g. *tablespoon*) most likely have their own lexical entries in addition to entries for their morphemic constituents (e.g. *table* and *spoon*). However, semantically transparent compounds do not have their own lexical entries and instead are decomposed into their parts prior to recognition.

The lexicon is organized in a number of ways (three are described here), and this organization affects how words are accessed during active recognition or production. First, the lexicon can be divided into 'closed-class' words and 'open-class' words. Closed-class words (also called 'function' words) comprise articles, prepositions, pronouns, and conjunctions. This class of words is so called because the membership changes little over time. In contrast, new open-class words can easily be added. Open-class words comprise nouns, verbs, adjectives, and adverbs and are often thought of as the 'content' words of a language.

Second, the lexicon is organized according to syntactic category. Lexical entries contain information about their syntactic category (for example, *cat* is a noun; *pet* can be a noun or a verb). In addition, many grammarians assume that these syntactic categories can be subdivided. For example, nouns can be divided into common nouns (nouns that can be preceded by a 'determiner' such as *a* or *the*) and proper nouns (names, places, events, etc.). Common nouns can be further divided into 'count nouns' (e.g. *pebble*) and 'mass nouns' (e.g. *water*)—count nouns can be used in the plural, but mass nouns ordinarily cannot. Thus, lexical entries contain information about minor categories (such as whether the entry is a count noun) as well as major ones (such as nouns and verbs).

Third, semantic relations among open-class words play an important role in determining the structure of the lexicon. Five of the most common relations are synonymy, antonymy, incompatibility, hyponymy, and meronymy. Synonymy is traditionally defined in terms of substitution. That is, two words are said to be 'synonyms' if they can be interchanged without affecting the truth value of the statements in which they appear (e.g. *sofa* and *couch*). It should be noted that synonyms can rarely be exchanged in every possible context; thus, synonymy is also thought of as a case of extreme semantic similarity. Words are said to be 'antonyms' if they are lexical opposites (e.g. *hot* and *cold*). 'Incompatibility' refers to a relation in which two words cannot be used simultaneously (e.g. an object can be either a *circle* or a *square*, but never both). 'Hyponymy' refers to the inclusion of one class in another (e.g. *apple* is a type of *fruit*). 'Meronymy' is a semantic relation that exists between words that denote parts and wholes (e.g. a *stem* is part of an *apple*).

By measuring how quickly and accurately people can recognize or name a word, psycholinguists have identified several variables that influence access to the mental lexicon. One of the most robust findings is that high-frequency words take less time to access than low-frequency words. However, this finding appears to hold only for open-class words. A second finding is that concrete words (e.g. *apple* or *cat*) are more readily accessed than abstract words (e.g. *freedom* or *democracy*). A third finding is that the ease of accessing a word is affected by whether a semantically similar word has been recently presented in the experiment. That is, a word (e.g. *doctor*) is accessed more easily if it has been preceded by a semantically associated word (e.g. *nurse*) than if it has been preceded by an unrelated word. A fourth finding indicates that lexical access is affected by how a word sounds. Researchers have shown that when people have difficulty retrieving a word (a situation known as the 'tip-of-the-tongue phenomenon'), they tend to retrieve words with similar sounds rather than words with similar meanings. Taken together, these findings demonstrate that the organization of the lexicon is very

flexible and that grammar, meaning, and pronunciation can all play a role in retrieving a word from the lexicon.

References

Aitchison, Jean. 1987. Words in the mind: an introduction to the mental lexicon. Oxford, UK: Blackwell.

Bresnan, Joan (ed.) 1982. The mental representation of grammatical relations. Cambridge, MA: MIT Press.

Cruse, D.A. 1986. Lexical semantics. Cambridge, UK: Cambridge University Press.

Jackson, Howard. 1988. Words and their meaning. London, UK: Longman.

Marconi, Diego. 1997. Lexical competence. Cambridge, MA: MIT Press.

Pustejovsky, James. 1996. The generative lexicon. Cambridge, MA: MIT Press.

Spencer, Andrew. 1991. Morphological theory: an introduction to word structure in generative grammar. Oxford, UK: Blackwell.

Stowell, Tim, and Eric Wehrli (eds.), 1992. Syntax and semantics, Vol. 26: syntax and the lexicon. New York: Academic Press.

CHRISTINA L. GAGNÉ

Lingua Franca

Although the term 'lingua franca' is best known as a common noun, meaning a language used between people who have no language in common, this entry is about the Lingua Franca in its proper noun meaning, i.e. the Italian-based contact language that was used throughout the Mediterranean area from the fifteenth to the nineteenth century. In this usage of the term, 'Franca' refers to the Franks, the word that was used for Europeans in the Arabic-speaking world of the Late Middle Ages. While 'Lingua Franca' literally means 'European language', its meaning was narrowed down to 'Romance-based pidgin'. Although Lingua Franca is sometimes referred to as Sabir, this may be confusing, since the latter is also used to refer to the French-based pidgin that came into existence (presumably on the basis of Lingua Franca) in North Africa in the late nineteenth century.

Although Lingua Franca is traditionally categorized as a pidgin language, there is some reason to qualify this classification. As was already observed by Schuchardt (1909), some of its linguistic features, such as the generalized use of the infinitive, suggest that Lingua Franca may perhaps be more accurately viewed as a form of Foreigner Talk. Other linguists (e.g. Minervini 1996) have claimed that it should rather be seen as a second-language variety of Italian. And since Italian and Spanish, the languages that formed the basis for Lingua Franca, were closely related dialects rather than separate languages five centuries ago, it might perhaps more appropriately be categorized as a koiné, i.e. the product of dialect convergence.

The Lingua Franca arose as a trade language in the context of the colonization of the Eastern Mediterranean by the Italian city-states of Venice and Genoa, whose dialects left a stamp on the language, for example, in the first and second singular pronouns *mi* and *ti*, which may well be of Venetian origin. Lingua Franca was diffused to the Western Mediterranean when the Genoese moved their commercial activities to the Iberian peninsula and North Africa, where it acquired an important role as a medium of intercultural communication in places such as Tripoli, Tunis, and, especially, Algiers. The presence of Turkish-speaking soldiers, Berber and Arabic-speaking Moors, and Christian slaves and renegades speaking a variety of European languages turned Algiers into a multilingual city *par excellence*. Due to the Spanish presence in North Africa, a number of Spanish words were incorporated, which resulted in Spanish–Italian doublets such as *fazer* (from Spanish) and *counchar* (from Italian), both meaning 'do' or 'make'. It is this presence of Spanish-derived words that forms the major difference between the western and eastern variety of Lingua Franca.

Since it is no longer spoken, all information on the language is restricted to written sources. Some of these are fictional works, such as poems and plays in which Oriental characters are depicted as speaking some sort of restructured Italian. The problem with this type of source, of course, is to decide whether they reliably reflect the language as it was actually spoken. The best known among these literary sources is Molière's (1670) *Le bourgeois gentilhomme*, which contains the famous lines *Se ti sabir, ti respondir. Se non sabir, tazir, tazir* 'If you know, answer. If you don't know, keep quiet'. Another author who has made extensive use of Lingua Franca for comical effect is the eighteenth-century Venetian playwright Carlo Goldoni. Although Schuchardt (1909) felt Goldoni's

Lingua Franca to be 'colorless', it is still worthwhile to consult Kahane and Kahane (1976) for a survey of its linguistic features.

A more reliable source of information can be found in nonfictional texts, such as historical and geographical works. The most extensive text of this type is a booklet entitled *Dictionnaire de la langue franque ou petit mauresque* (henceforth *Dict.*), published anonymously in Marseilles in 1830. It is a 112-page language primer, containing a wordlist and eight short dialogues, meant to provide the French military with a basic knowledge of the language when they embarked upon their invasion of Algiers. Another important documentary source is Haedo's (1612) *Topographia e historia general de Argel*, a geography and history of Algiers, partly based on an eyewitness account. Apart from the Lingua Franca fragments, the book also contains some acute observations on the language situation in Algiers at the turn of the sixteenth century. Haedo's remarks leave no doubt that the Lingua Franca played an important role as a medium of interethnic communication among Moors, Turks, and Europeans in that city.

As to its main linguistic features, the following may be mentioned. While the lexicon is mainly derived from Italian and Spanish, some words have been taken from Arabic, such as *taybo* 'good' and *marfuz* 'bad', and from Turkish, such as *Yoldach* 'Turk'. The morphology is more elaborate than is usually the case with pidgin languages. Agreement between the head and its modifier(s) is quite common, especially with regard to gender, as in *barbero bono* '(a) good doctor' and *bona bastonada* '(a) severe beating' (both from Haedo). Although verbal inflection is heavily reduced when compared to the lexifier languages, we find occasional use of the past participle to express past tense, as in *mi mirato iéri* 'I saw him yesterday' (lit. I seen (him) yesterday) (*Dict.*) The only other verbal form is the infinitive, sometimes combined with an adverb such as *bisognio* 'necessary' to express future or irrealis, as in *cosa bisognio counchar?* 'What shall we do?' or 'What should be done?' (lit. What necessary do?) (*Dict.*). Although reduplication is generally rare in pidgins (as opposed to creoles), it occurs quite often in the Lingua Franca, as in *mucho mucho* 'very much' (Haedo) and *siéme siéme* 'together' (*Dict.*).

The basic word order is S(ubject)-V(erb)-O(bject), although VS order is found as well, e.g. in sentences containing a focused constituent and in subordinate clauses. In contrast to many other pidgins, the copula is usually overtly expressed (by *(e)star*), as in *qouesto star véro* 'That is true' (*Dict.*). Pronominal subjects, especially third singular, often remain unexpressed, a feature that may be attributed to Italian and Spanish influence. Direct as well as indirect pronominal objects are often marked by the preposition *per*, as in *mi star contento mirar per ti* 'I'm happy to see you' (lit. I happy see PER you) (*Dict.*). Complex sentences are quite common, with the subordinate clause being attached to the matrix clause either by a conjunction, a relativizer, or a complementizer, or without any overt linking element at all. The latter occurs especially with object sentences, as in *mi pensar star meïo* 'I think it's better' (*Dict.*).

References

Anonymous. 1830. Dictionnaire de la langue franque ou petit mauresque. Marseilles: n.p.

Arends, Jacques. 1998. A bibliography of Lingua Franca. The Carrier Pidgin 26. 4–5; 33–5.

Arends, Jacques. 1999. Lingua Franca en de 'Europese' creooltalen (Lingua Franca and the 'European' creoles). Gramma-TTT 7. 173–90.

Cifoletti, Guido. 1989. La Lingua Franca Mediterranea. Padova: Unipress.

Haedo, Diego de. 1612. Topographia e historia general de Argel, Valladolid; republished (3 vols.), Madrid: Sociedad de Bibliófilos Españoles, 1927–1929.

Kahane, Henry, and Renée Kahane. 1976. 'Lingua Franca': the story of a term. Romance Philology 30. 25–41.

Metzeltin, Michael. 1996. Lingua Franca. Kontaktlinguistik — Contact linguistics — Linguistique de contact, Vol. 1, ed. by Hans Goebl et al., 554–8. Berlin and New York: De Gruyter.

Metzeltin, Michael 1998. Die linguae francae des Mittelmeers (The lingua francas of the Mediterranean) Lexikon der romanistischen Linguistik, Vol. 8, ed. by Günther Holtus, Michael Metzeltin and Christian Schmitt, 601–10. Tübingen: Niemeyer.

Minervini, Laura. 1996. La Lingua Franca Mediterranea: Plurilinguismo, mistlilinguismo, pidginizzazione sulle coste del Mediterraneo tra tardo medioevo e prima età moderna (The Mediterranean Lingua Franca: Multilingualism, language mixing, and pidginization along the Mediterranean coast in the Late Middle Ages and the Early Modern Period). Medioevo Romanzo 20. 231–301.

Molière, Jean-Baptiste Poquelin de. 1670. Le bourgeois gentilhomme. Molière, Oeuvres complètes, ed. by Robert Jouanny. Paris: Bordas, 1993.

Schuchardt, Hugo 1909. Die Lingua Franca. Zeitschrift für Romanische Philologie 33; translated as The Lingua Franca by Glenn Gilbert. Pidgin and creole languages: selected essays by Hugo. Schuchardt, ed. by Glenn Gilbert, 441–61. Cambridge: Cambridge University Press, 1980.

Whinnom, Keith. 1977. Lingua Franca: historical problems. in Pidgin and creole linguistics, ed. by Albert Valdman, 295–310. Bloomington: Indiana University Press.

The entire Dictionnaire de la langue franque along with much information on the Lingua Franca can be found at <http://www.uwm.edu/~corre/franca/go.html>

JACQUES ARENDS

See also **Language: Contact—Overview; Pidgins and Creoles**

Literacy

The term literacy is most commonly interpreted as being the ability to read and write. However, focusing on decoding and encoding leads to meaningless techniques that obscure a broader understanding of this complex cultural phenomenon. Ever since the first western postwar warnings about the 'literacy crisis' and the widespread 'illiteracy' among schooled people of industrialized countries, questions have arisen that challenge the validity of this focus.

Critical debates involving social history, sociolinguistics, educational, and cultural studies have led to a specific field of research that questions the psycholinguistic view of literacy as a unified, abstract, and universal human cognitive capacity relating to the use of written materials, and describes the variety of epistemological frameworks of the current research on approaches and attitudes toward literacy.

The challenge, therefore, has been to provide a broader understanding of the nature and consequences of writing and reading as political, social, mental, physical, and linguistic actions, rather than merely individual standardized technomechanical skills.

The greatest concern in literacy studies still focuses on the fact that definitions, values, and functions ascribed to literacy are totally ambiguous unless they are viewed in specific sociohistorical contexts. The crucial question is what to consider as literacy in a given community or in a given historical epoch. From a pragmatic and ethnographic standpoint, for instance, someone who simply signs legal documents may be considered literate in rural areas, whereas someone with the skills required to decode written text, but whose ability to make sense of that text or to act upon it is not developed, is seen as illiterate in postindustrial communities. From the same viewpoint, shop floor workers who can read, follow written instructions, and write occasional short notes may be considered illiterate in a technological and computer-based work context, in which information must be fed into and taken out of computer terminals.

As these examples clearly demonstrate, recognizing and manipulating patterns in print and linking these patterns to oral language may not suffice to perform successfully in a given sociohistorical or socioprofessional context. Moreover, standardized test scores that quantify literate competence over time and across space may conceal extremely different and continually changing sociocultural realities. Hence, criteria for both literacy and illiteracy should be based on degrees in a continuum rather than on categorical distinctions and universal quantitative rates, rendering the traditional general distinctions between literacy and illiteracy, and between esthetically oriented literacy and functional or instrumental literacy useless. Any meaningful definition of literacy or illiteracy must refer to a particular sociocultural context.

A further implication of this relativistic perspective is that literacy as a subject of scientific inquiry concerns both the myriad social practices of which the activities of reading and writing are an integral part and the multitude of sociocultural concepts, values and functions of what is considered a literate performance in a given situation. The focus, therefore, is on the contextual use of language and not on abstract, context-free structures or on written products of communication. Moreover, if reading and writing are essential elements of cultural practices, i.e. cultural ways of doing things, and make sense within groups and communities, they must, perforce, be seen as deeply interconnected with and inseparable from each other.

By focusing on everyday uses of reading and writing in local communities, detailed ethnographic studies have played an important role in the investigation and description of the variety of events and communication practices that comprise the so-called 'ecology' of literacy practices. Identifying different forms of literacy in different domains of life, such as computer literacy, academic literacy, workplace literacy, etc., and developing the notion of hidden forms of literacy (in contrast to visible or dominant forms of literacy), ethnographic studies have also been the basis for insightful discussions about the strict divide traditionally drawn between oral and literate cultures and subcultures, as well as about the relationship between local forms of literacy and formal training. Some of the most telling findings of these studies concern the complementary function of oral and written forms in literate communication practices; the processes of informal learning in the acquisition of literacy practices; the discrepancy between what counts as literacy in school and in other mainstream institutions; and the pervasive role of configurations of power and social knowledge structures in shaping institutional and non-institutional literacy practices.

Another key issue in literacy studies has been the study of the consequences of literacy on language use

and on intellectual and social life. In general terms, the social view of literacy extended the notion of text and of written language usage conceived by theoretical linguistic traditions, demolishing an array of mythologies inherited from philosophical and scientific discourses about the intrinsic value of writing due to the assumed relationship between scriptural order and reason, and between literate competencies and cognitive, technoscientific, and socioeconomic development.

The link between the logic of alphabetic writing and rationality or logical reasoning was made in the 1960s and 1970s by English-speaking scholars, who integrated and extended the ideas about the cognitive and social impact of literacy. These scholars attempted to demonstrate that a 'literate mind' is cognitively distinct from a nonliterate one, since it is highly skilled in thinking and speaking abstractly, clearly, and reasonably. Accordingly, it was also asserted that a literate society possesses a distinctive form of organization and is highly skilled in technoscientific development.

Their historical, anthropological, and psychological studies about different, mainly nonindustrialized, culture has led many scholars, concerned with how the acquisition of literacy affects individuals and societies, to corroborate and reinforce some traditional assumptions about literacy as a skill-specific competence, or a cognitive condition, and about the 'great divide' between speaking and writing and between oral and literate cultures. Another implication of these assumptions is that literacy is identified as a neutral and powerful technology, which positively transforms human mental capacities and social behavior. Accordingly, its appropriation by individuals and societies is primarily a cognitive and technomethodological problem. As a result, illiteracy is identified as a cognitive handicap on both individual and collective levels, or a technomethodological insufficiency at an institutional level (for instance, the use of inappropriate equipment, sites, and pedagogical materials). Transferred to communities, therefore, these assumptions are simplistically constraining, reducing the idea of literacy by ignoring the fundamental issue of the sociopolitical regulation of access to literacy, which is associated with social inequalities of power and control.

In contrast to this perspective, and drawing on comparative socioanthropological studies, critical studies on literacy have found much evidence that different forms and levels of literacy are also shaped by social stratification and control: even nominally literate contemporary societies show institution-regulated and often pyramidal distributions of competence in basic formal and pragmatic conventions of writing. The resources of literacy to which people have access in households and communities are fundamentally unequal. Yet, as many cross-cultural and cross-historical comparisons within context-specific settings have demonstrated, what is naturalized or routinized by influential social groups is not equally acquired and used by an entire population, even with compulsory mass literacy training, which explicitly or implicitly embodies mainstream ideological and normative agendas, entailing very different consequences within particular cultures and subcultures. These studies also found that literacy practices had been historically embodied within religious, ethical, and ideological issues.

Competing perspectives of the social and cultural values and effects of literacy on the daily life of modern societies were also drawn from the prescriptive agendas of postwar western national institutions and international organizations such as UNESCO. Since the 1970s, the link between these agendas has been a strong commitment for the establishment and expansion of a standardized literate competence, as a basic human right to be achieved both in industrialized and nonindustrialized areas around the world. Following a line of reasoning and discussion that builds on commonplace literacy mythologies intertwined with arguments for competence in personal self-improvement, citizenship, and social mobility in modern nations, UNESCO's studies and campaigns, and many national educational policies inspired by them, have reinforced conventional assumptions about the 'civilizing' effects of literacy, while encouraging heated debates over the social purposes and political potential of literacy.

The studies and campaigns developed over the past few decades have been based on assumptions about the usefulness and value of universal literacy training in any social context, and have been so devised that illiteracy and the aforementioned "literacy crisis" have become a strategic governmental issue, triggering a number of local and national campaigns and massive financial investments in many nations around the world. These campaigns aimed to 'eradicate' illiteracy in the world before the year 2000. The analogy between illiteracy and disease has been a recurrent theme, widely disseminated through official texts and images.

It is generally recognized that the set of values and beliefs that underpinned institutional public discourses about literacy as a key element in the production of progressive mentalities and productive individuals for capitalist economies and modern nation-state building relates to the concepts of progress, development, and emancipation inherited from post-Enlightenment liberal social theories. Included here are beliefs in the liberating and revolutionary consequences of literacy, such as democracy and political participation.

Concomitant to such official views and guidelines, however, significant alternative viewpoints have been brought into public and educational discussions. Drawing on critical literacy studies and on a Marxist-oriented understanding of emancipative

education, many researchers have successfully demonstrated that the acquisition of literacy may entail both negative and positive aspects, and that it is necessarily associated with ideological and political issues. Accordingly, they have tried to demystify the fallacious equation of literacy with progress and social mobility, as well as the official rhetoric and current school practices that deliberately obscure the range of social, economic, and political forces that inform literacy practices and formal education. They claim that by promoting homogenizing and manipulative socialization, the acquisition of literacy has become a disempowering process (in contrast to the so-called 'empowering literacy'), which implies passivity, contributes nothing to personal improvement, and promotes social exclusion. Additionally, if literacy is often present in unequal social structures, this inequality should be the most important motivator for a reevaluation of the conceptions, functions, and uses to be ascribed to literacy within a given sociocultural context.

Having discarded a view of literacy as a context-neutral, skill-specific competence, works on literacy acquisition and language education have also challenged linguistic assumptions about the neutrality and transparency of writing as a modality of language usage. Consequently, they have also questioned school conceptions of literacy that promote the acquisition of strict instrumental skills, passive attitudes, and prescribed behavior, as well as the traditional hierarchization of literacy practices, textual genres, and linguistic modalities. In literacy training, emphasis has been placed on the social and political dimensions of the differences between different types of literacy, or different forms of literacy and, more specifically, on the whole issue of inequality between them.

In light of the different perspectives on literacy, and emphasizing the ideological dimension of any understanding of it, current research in this area focuses on the variety of written materials and communication practices of everyday life in different sociocultural networks and communities, in an attempt to throw light on the various forms of influence they may exert on the economic and political exclusion, or inclusion, of any given contemporary social group. Concomitant to this fundamental questioning is the attempt to promote a profound rethinking of the empowering properties underlying contemporary institutional and noninstitutional literacy practices. Needless to say, any and all new commitments in this field must take into account educational issues and the fact that literate competencies in postindustrial societies tend to be transformed into a set of available consumer objects like, for instance, reading training within highly structured and standardized interpretive schemata that fit in perfectly with standardized tests and performances.

References

Barton, David, and Ivanic, Roz (eds.) 1991. Writing in the community. Newbury Park, London and New Delhi: Sage Publications.

Freire, Paulo. 1970. The adult literacy process as cultural action for freedom. Harvard Educational Review 48.

Freire, Paulo. 1973. The pedagogy of the opressed. New York: Herder and Herder.

Goody, Jack. 1979. Domestication of the savage mind. Cambridge: Cambridge University Press.

Goody, Jack. 1986. The logic of writing and the organization of society. Cambridge: Cambridge University Press.

Graff, Harvey J. 1979. The literacy myth: literacy and social structure in the 19th century city. New York: Academic Press.

Graff, Harvey J. 1987. The labyrinths of literacy: reflections on literacy past and present. London: Falmer Press.

Heath, Shirley B. 1983. Ways with words: language, life, and work in communities and classrooms. Cambridge: Cambridge University Press.

Olson, David R. 1977. From utterance to text: the bias of language in speech and writing. Harvard Educational Review 47. 257–81.

Olson, David R. 1986. Learning to mean what you say: toward a psychology of literacy. Literacy, society and schooling, ed. by Suzanne de Castell, Allan Luke and Kieran Egan. Cambridge: Cambridge University Press.

Scribner, Sylvia, and Michael. Cole. 1981. The psychologie of literacy. Cambridge, MA: Harvard University Press.

Street, Brian V. 1984. Literacy in theory and practice. Cambridge, New York and Melbourne: Cambridge University Press.

INÊS SIGNORINI

Localization of Linguistic Information

The subject of localization is the relationship of function with neuronal structure. The main issues of localization are the definition of the concept of a function, the description of the anatomical structures necessary to carry out such a function in the normal brain, and the description of the duty of a certain area of the

brain. Generally, one may differentiate between the localizability of simple functions such as moving a leg and complex functions such as language processing. Simple functions have the greatest anatomical and structural correlation whereas complex functions apparently have relatively widespread cortical and subcortical localization.

Before the nineteenth century, the common theory of brain function, which was termed "holism", assumed that the brain simply was not a structure made up of discrete independent centers, each specialized for different functions, but must be looked upon as a single working unit. However, at the beginning of the nineteenth century, scientists began to assign function to certain neuronal structures and started to favor the theory of localization of function within the brain. The first localization of linguistic function to a specific region and hemisphere of the brain that became widely accepted is usually attributed to the French surgeon Pierre Paul Broca (1824–1880) and the German neurologist Carl Wernicke (1848–1905). They examined the brains of individuals who had suffered a stroke, become aphasic, and later died. Based on postmortem correlation, Broca suggested that the impaired articulation of speech was due to a lesion within the left inferior frontal gyrus of the brain (Broca's area). Wernicke correlated lesions in the left superior temporal gyrus (Wernicke's area) with disorders in language comprehension (Figure 1). Wernicke, who was neither a pure localizationist nor a pure holist, also developed an elaborate model of language processing, where he proposed that only basic perceptual and motor activities are localized to single cortical areas and interconnections between these functional sites make more complex intellectual functions possible. He stated that different components of a single behavior are processed in different regions of the brain and thus advanced the first evidence for the idea of distributed processing and connectionism. In the first part of the twentieth century, the idea of functional segregation again fell into disrepute, proposing that higher cognitive abilities depended on the function of the brain as a whole. However, at this time most neuroscientists were willing to accept some sort of cortical localization for sensory and motor function. In contrast, the situation concerning memory and language was still an active battleground for the opponents and defenders of localization. Until the 1960s, most of the information about the localization of linguistic function was based primarily on patients with brain lesions. In this case, the language deficits resulting from brain injuries have been compared to the areas of the brain which became lesioned. Even though the lesion can be accurately located, the function that is examined after injury does not reflect the simple equation of the

normal function missing, but it represents a new state of reorganization of the brain. In the 1960s the American neurologist Norman Geschwind (1926–1984) refined Wernicke's model of language processing, and this so-called Wernicke–Geschwind model still forms the basis of current investigations on normal and disturbed language function. Based on data from a large number of patients and a thorough understanding of cortical connectivity, Geschwind argued that regions of the parietal, temporal, and frontal brain lobes were critically involved in human linguistic capacities. This model holds that the comprehension and formulation of language are dependent on Wernicke's area, after which the information is transmitted over the arcuate fasciculus to Broca's area where it can be prepared for articulation.

The further expansion and improvement of this model of language processing within the brain has been made possible by the application of advanced neurophysiological methods. Electrophysiological studies using electric or magnetic stimulation and mapping techniques provided a way to delineate the language areas of the cortex prior to neurosurgical removal of brain tissue. These intraoperative mapping techniques have contributed interesting localization information, and provided evidence that a large region of the left hemisphere is clearly involved in language production and comprehension and that language localization varies from patient to patient. Modern functional neuroimaging techniques such as positron emission tomography (PET), established in the early 1980s, and functional magnetic resonance imaging (fMRI), established in the 1990s, represent a major step toward better understanding the localization of linguistic function. Both techniques are imaging procedures that visualize local changes in cerebral blood flow and metabolism that accompany language processing in normal subjects. Contemporary studies are offering new data on what brain areas might actually do and how they might contribute to a network of neuronal structures that collectively participate in language processing. The results obtained by these neurophysiological methods support the notion that besides the classical Broca's and Wernicke's areas, several additional distributed cortical and subcortical neuronal structures of both hemispheres clearly make a significant contribution to language function (see Figure 1).

Most of the areas consistently activated by linguistic processing lie within the left temporal, parietal, and frontal brain lobes (Figure 1) and partly in right hemispheric homolog areas. Extended areas in the left temporal cortex are engaged in word retrieval, naming, morphosyntactic processing, parsing, syntactic comprehension and semantic analysis and also in the articulation of speech. Moreover, several areas in the
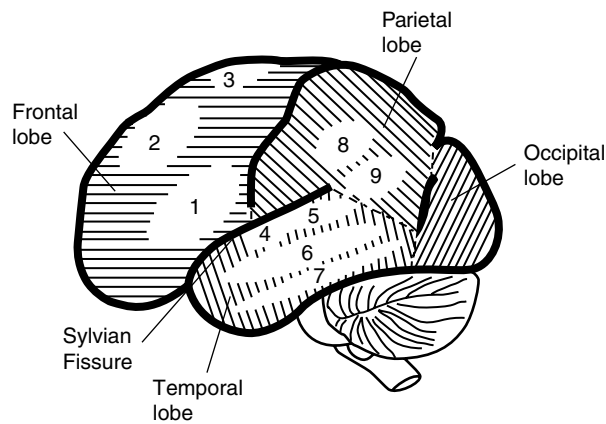
Figure 1. Left hemispheric cortical brain areas related to auditory linguistic processing. Broca's area and parts of middle frontal gyrus (1); dorsolateral prefrontal cortex and parts of superior frontal gyrus (2); supplementary motor area (3); primary auditory cortex (4); Wernicke's area (5); parts of middle temporal gyrus (6); parts of inferotemporal gyrus (7); supramarginal gyrus (8); angular gyrus (9).

left frontal cortex including Broca's area are found to be concerned with word retrieval, verbal working memory, syntactic processing, semantic encoding and retrieval, and also articulation, comprehension, and global attentional or executive functions. Right hemispheric cortical areas that correspond to Broca's and Wernicke's areas in the left hemisphere are associated with prosodic elements that impart additional meaning to verbal communication and additional linguistic functions that have not been clearly defined up to now. Moreover, the gyrus cinguli, the cerebellum, and subcortical regions like the basal ganglia and especially the thalamus, which is presumably a lexical–semantic interface, may not only play supportive roles during linguistic processing.

The fact that there are so many new brain regions emerging in modern lesion and functional imaging studies of language suggests that the classical Wernicke–Geschwind model, although useful for so many, years has now been seen to be oversimplified. Areas all over the brain are recruited for language processing, some are involved in lexical retrieval, some in grammatical processing, some in the production of speech, and some in attention and memory. These new findings are still too fresh for any overarching theories that might explain how these areas interact. However, there is a definite need for hybrid models of language processing. Language is both localized and distributed, which means that specific language processing operations are carried out in particular brain locations organized in a distributed fashion.

References

Broca, Paul. 1861. Remarques sur le Siége de la Faculté du Langage Articulé, Suivies d´une Observation d´aphémie (perte de la parole). *Bulletins de la Societé Anatomique* 36.

Brown, Colin M., and Paul Hagoort (eds.) 1999. *The neurocognition of language.* Oxford: Oxford University Press.

Finger, Stanley. 1994. *Origins of neuroscience.* New York: Oxford University Press.

Gazzaniga, Michael S. (ed.) 2000. *The new cognitive neurosciences.* Cambridge, MA: MIT Press.

Geschwind, Norman. 1974. *Selected papers on language and the brain.* Dordrecht: Reidel.

Kertesz, Andrew, (ed.) 1994. *Localization and neuroimaging in neuropsychology.* San Diego: Academic Press.

Kosslyn, Stephen M., and Richard A. Andersen, (eds.) 1992. *Frontiers in cognitive neuroscience.* Cambridge, MA: MIT Press.

Lashley, Karl, S. 1929. *Brain mechanisms and intelligence.* Chicago: University of Chicago Press.

Ojemann, George A. 1991. Cortical organization of language. *Journal of Neuroscience* 11.

Stemmer, Brigitte, and Harry A. Whitaker (eds.) 1998. *Handbook of neurolinguistics.* San Diego: Academic Press.

Uttal, William R. 2001, *The new phrenology: the limits of localizing cognitive processes in the brain,* Cambridge, MA: MIT Press.

Wernicke, Carl, 1874. *Der aphasische Symptomenkomplex: eine psychologische Studie auf anatomischer Basis*, Breslau: Cohn und Weigert; as *Wernicke's works on aphasia: a sourcebook and review*, translated by Gertrude H. Eggert. The Hague: Mouton, 1977.

SABINE WEISS

Long-Distance Dependency

The phenomenon of long-distance dependency occurs in languages such as Latin, Russian, Japanese, Korean, or Persian, in which the order of the major constituents of the clause is variable. These languages, known as 'free word order' languages, sharply contrast with languages such as English, which exhibit a relatively fixed word order. In the following German (1) and Japanese (2) examples, for instance, the direct object

may either precede or follow the subject, a possibility that is not available in English:

(1) ... weil niemand **dieses Buch** gekauft hat
... because nobody this book bought has
'... because nobody has bought this book.'
... weil **dieses Buch** niemand gekauft hat

(2) John-ga Mary-ni piza-o ageta
John-NOMINATIVE Mary-DATIVE pizza-ACCUSATIVE gave
'John gave Mary pizza.'
Piza-o John-ga Mary-ni ageta

Ross (1967, 1986) coined the term *scrambling* to refer to this variable word order and proposed a universal rule to account for it; the rule stated that two adjacent constituents can be permuted if they are constituents of the same minimal clause. It follows from Ross's definition of the scrambling rule that this phenomenon is, as in the examples above, clause-bound. However, some languages additionally exhibit so-called long-distance scrambling, which involves the movement of a constituent across a finite clause boundary. In the following Russian sentence, the noun phrase 'Petrov' has scrambled to a higher clause out of the *that*-clause between brackets:

(3) On skazal čto **Petrov**   stranno [čto ------ nam
he said that Petrovit-NOM is odd   that   us.

pomogal]
helped

It is claimed in the literature that scrambling is not a unitary phenomenon crosslinguistically. There are languages, English among them, that do not allow scrambling, languages that only have scrambling within the verb phrase (e.g. Hungarian), or clause-internally (e.g. German), and languages with a more flexible word order in which a constituent may appear outside the clause it belongs to. These free word order languages, on the other hand, also differ as regards the long-distance scrambling operations they allow, since this phenomenon is subject to language-specific restrictions. In Japanese, for example, most direct objects can undergo long-distance scrambling, but when subjects are scrambled, the sentence turns out to be ungrammatical:

(4) **Sono hon-o**   John-ga   Bill-ni [Mary-ga-----motteiru
that book-ACC John-NOM Bill-DAT Mary-NOM   have

to] itta
that said

(5) * **Sono hon-ga**   John-ga [------yoku ureteiru to]
that book-NOM John-NOM   well sell   that

omotteiru
think
* 'That book, John thinks (it) sells well.'

In contrast to Japanese, Serbo-Croatian, in which word order is much more flexible, allows both objects and subjects to be scrambled to the matrix clause:

(6) Verujem **novac** [da Petar ------daje Milanu]
I-believe money that Petar   gives to-Milan
'I believe that Peter gives money to Milan.'

(7) Verujem **Petar** [da ------novac daje Milanu]
I-believe Petar that money gives to-Milan

Crosslinguistic investigations have also shown that there are differences with respect to the syntactic positions in which scrambled elements can be located within the matrix clause. In Persian, a constituent undergoing long-distance scrambling can be placed both between the matrix subject and the verb (8) and in the clause-initial position (9). On the contrary, it seems that in Japanese this process is somehow restricted to the clause-initial position, as shown above in (4), since long-distance scrambling to the intermediate position within the matrix clause does not always result in a fully grammatical sentence (10):

(8) Ali **een ketab-ra** fekr-mikone [ke Mehry ------ be
Ali this book   think   that Mehry   to

Hassan dad]
Hassan give

'Ali thinks that Mary gave this book to Hassan.'

(9) **Een ketab-ra** Ali fekr-mikone [ke Mehry ------ be
Hassan dad]

(10) ??John-ga **sono hon-o** minna-ni [Hanako-ga ----
John-NOM thatbook-ACC everyone-DAT Hanako-NOM

yonda to] itta
read that said
?? 'John, that book, said to everyone that Hanako read.'

Another remarkable feature of long-distance scrambling is that in some languages this grammatical process can occur more than once within a single clause. The following Korean sentence, in which scrambling has been applied to both direct and indirect objects, illustrates this possibility:

(11) **Kulim-ul**$_i$ **Yenghi-eykey**$_j$ Chelswu-ka [ai-ka ------$_j$
Picture-ACC Yenghi-DAT   Chelswu-NOM child-NOM

------$_i$ cwu-ess-ta-ko] sayngkakha-n-ta
give   think

'Chelswu thought that the child gave the picture to Yenghi.'

The phenomenon of long-distance scrambling has recently received a great deal of attention. However, there is no consensus at the moment about certain descriptive and theoretical aspects concerning this

grammatical process. German, for example, is traditionally analyzed as a language that does not have long-distance scrambling. In fact, a finite clause boundary may never be crossed by a scrambled constituent in this language:

(12) *...weil Hans **den Wagen** versprochen hat [dass
because Hans the car promised has that

er ------ reparieren würde]
he repair would

'... because Hans has promised that he would repair the car.'

Nevertheless, German exhibits a structure, sometimes referred to as the 'Third Construction', which closely resembles long-distance scrambling: there are a special class of verbs—*versprochen* 'try', *versuchen* 'try', *beginnen* 'begin' or *hoffen* 'hope', among others—that allow scrambling to the matrix clause out of their infinitival complements:

(13) ... weil Hans **den Wagen** versprochen hat [------ zu reparieren]

... because Hans the car promised has to repair
'... because Hans has promised to repair the car.'

For some authors, this construction counts as an instance of long-distance scrambling and therefore they classify German together with Japanese or Russian as a long-distance scrambling language.

Turning finally to the theoretical side of this phenomenon, many recent generative approaches to scrambling assume that the alternate word order arrangements in a language are the result of a syntactic movement applied to a single basic structure. Yet there is no agreement on which kind of movement is involved: some authors argue that long-distance scrambling is a case of NP-movement—which also gives rise to passive sentences in English—while others analyze it as an instance of *wh*-movement—which forms English *wh*-questions. An alternative theory of scrambling in generative literature rejects both any type of movement and the existence of a single basic structure. Under this assumption, the alternate constituent orders are all generated as basic structures without any derivational relationship between them.

References

Corver, Norbert, and Henk van Riemsdijk (eds.) 1994. Studies on scrambling. Movement and non-movement approaches to free word-order phenomena. Berlin and New York: Mouton de Gruyter.

Culicover, Peter W. 1996. On distinguishing movements. Linguistic Inquiry 27.

Fukui, Naoki. 1993. Parameters and optionality. Linguistic Inqury 24.

Grewendorf, Günther, and Joachim Sabel. 1994. Long scrambling and incorporation. Linguistic Inquiry 25.

Grewendorf, Günther, and Joachim Sabel. 1999. Scrambling in German and Japanese: adjunction versus multiple specifiers. Natural Language and Linguistic Theory 17.

Grewendorf, Günther, and Wolfgang Sternefeld (eds.) 1990. Scrambling and barriers. Amsterdam and Philadelphia: John Benjamins.

Miyagawa, Shigeru. 1997. Against optional scrambling. Linguistic Inquiry 28.

Müller, Gereon, and Wolfgang Sternefeld. 1993. Improper movement and unambiguous binding. Linguistic Inquiry 24.

Nemoto, Naoko. 1995. Scrambling in Japanese, AGRoP, and economy of derivation. Lingua 97

Ross, John R. 1967. Constraints on variables in syntax. Ph.D. Dissertation. Massachusetts Institute of Technology; as Infinite Syntax! Norwood, NJ: Ablex, 1986.

LUISA GONZÁLEZ-ROMERO

Long-Range Comparison

Long-Range Comparison (LRC) is an area of linguistics that is open to extralinguistic concerns. This has caused it to be popularized outside narrow circles of specialists. The idea that all human languages and language families might be ultimately relatable has probably been current since people started speculating about human language. There are numerous attestations of such speculations, often linked to the tale of the Tower of Babel (Genesis 11:1–9), with various refinements. LRC is an essential link in this process.

In brief detail, LRC is aimed at establishing putative relationships among or between large linguistic groupings, and pushing back the frontiers of linguistic reconstruction. Although he was not the first to note the relevant parallels, Modern LRC is often dated from Holger Pedersen's coining of the term Nostratic (from the Latin "ours"), to include Indo-European, Uralic, Altaic, and Semitic. In the 1960s Illič-Svityč compiled the first comparative dictionary of Nostratic. He added Kartvelian and Dravidian to Pedersen's grouping.

Many LRC theories apply to whole continents, groups of continents, and even the whole of the Earth (proto-World). One prominent figure in this field is Joseph Greenberg, who has a long-standing involvement in such scholarship. Greenberg's earlier work on African linguistic relationships is generally accepted by linguists, although his later work on Eurasiatic (partly, but not totally, covering the same areas as Illi -Svity's Nostratic) and Amerind (relating nearly all the languages of North and South America) has come in for much greater criticism, especially the latter; see especially Matisoff (1990), Dixon (1997), Lass (1997:160–71), and the detailed treatments in McMahon and McMahon (1995), Watkins (1997:3), Trask (1999). Another such proposal is Dene-Caucasian, combining North Caucasian, Yeniseian, Sino-Tibetan, and Na-Dene, which has also been the subject of criticism. Greenberg's disciple Merritt Ruhlen has been more daring in his proposal of reconstructing a single origin for all human languages (1994). Some scholars (e.g. Dixon 1997:43) even lament that there is a certain popular appeal to the concept of all human languages having a single ancestor ("Proto-World"). Dixon cites numerous references to areal factors, typology, etc., to set up an alternative view of linguistic development, which he dubs "punctuated equilibrium". He continues its development, and adds more criticism of LRC, in Dixon and Aikhenwald (1999).

In some cases, LRC work is hampered by the fact that many of the putative families on which the respective superfamilies are based are themselves on rather shaky ground, or their exact relationships are still being worked out. Originally, Turkic, Mongolian, and Tungus (some scholars would also include Korean and Japanese) were grouped together within the Altaic family, which often figures in various LRC theories, but recently the very status of Altaic itself has been seriously called into question. Meanwhile, Semitic has been found to be part of a much wider family conventionally dubbed Afro-Asiatic. Afro-Asiatic in turn is often compared to Indo-European; see especially the work of Bomhard (1984, 1996). The status of Afro-Asiatic within the field of LRC has also been open to some discussion: many Nostraticists would class it under Nostratic, whereas other lines of research, e.g. Greenberg's Eurasiatic theory, would group it as a family equal in status to Nostratic with a more remote level of relationship.

The existence of typological similarities between the languages and language families compared often makes LRC more difficult. Using typological parallels in genetic reconstruction is equivalent to e.g. suggesting that whales are fish based on similarities in form. One excellent example is provided by the emergence of the laryngeal theory in Indo-European studies, boosted by the discovery and interpretation of Hittite. The existence of similar segments in Semitic suggested some sort of special relationship between Indo-European and Semitic. The apparent parallels in ablaut between Indo-European and Kartvelian provide another such example. In the last century, it was believed that Uralic and Altaic formed a superfamily (still referred to as Ural-Altaic in popular literature) based primarily on typological features common to both: progressive vowel harmony and agglutination. Other examples of LRC have often been based on similarities in word order, which is often subject to rapid change.

Much of the controversy, and many of the problems, may be illustrated by citing the development of research into the history of Basque. As a language isolate, Basque is of particular interest here, as it is not included in many proposals for superfamilies which do include almost everything from Kamchatka to the Sahara. However, this has not stopped some scholars from trying. Trask (1996:358-429) provides an overview, including extensive discussion, sometimes a little harsh, on failed attempts to fit Basque into various long-range proposals.

Russian scholars, such as the late V.M. Illič-Svityč, Dolgopolsky, Shevoroshkin, and Starostin, have contributed a great deal to recent work in LRC. Much of the material is in Russian, but see Shevoroshkin and Markey (1986).

Although it often appears that LRC concentrates on lexemes, it should be noted that Illič-Svityč listed pages of useful suggestions on pronouns, affixes, morphemes, etc. Nor does more recent work on Nostratic lack a discussion of morphology; cf. Bomhard (1996:90–101) and many of the articles in Salmons and Joseph (1998). To take one example, the use of a dental stop to form some sort of preterite over wide swathes of the Nostratic area, also extending to Japanese, certainly deserves some sort of discussion, even if much LRC work appears exaggerated.

A further factor that is assuming increasing importance and shedding light on purely linguistic issues is the ongoing work on human genetics. This work is still in its early stages, but is already yielding interesting results, e.g. possible support for Greenberg's Amerind theory, see Cavalli-Sforza, et al. (1994).

References

Baldi, Philip. 1990. Linguistic change and reconstruction methodology Berlin and New York: Mouton de Gruyter.

Bomhard, Allan R. 1984. Towards Proto-Nostratic: a new approach. Amsterdam and Philadelphia: John Benjamins.

———. 1996. Indo-European and the Nostratic hypothesis. Charleston, SC: Signum.

Cavalli-Sforza Luca L., Paolo Menozzi, and Alberto Piazza 1994 The history and geography of human genes. Princeton: Princeton University Press.

Dixon, Robert M. W. 1997 The rise and fall of languages. Cambridge: Cambridge University Press.

Dixon, Robert M. W., and Alexandra Y., Aikhenwald, 1999. The Amazonian languages. Cambridge: Cambridge University Press.

Greenberg, Joseph 1987. Language in the Americas Stanford: Stanford University Press.

Lass, Roger 1997. Historical linguistics and linguistic change. Cambridge: Cambridge University Press.

Matisoff, James 1990. On megalocomparison Language 66(1). 106–20.

McMahon, April M.S., and Robert McMahon 1995. Linguistics, genetics, and archaeology: internal and external evidence in the Amerind controversy". Transactions of the Philological Society 93(2).125–225.

Pedersen, Holger 1967. The discovery of language: linguistic science in the nineteenth century, translated by John Webster Spargo.Bloomington: Indiana University Press.

Renfrew, Colin, and Nettle, Daniel, (eds.) 1999. Nostratic: examining a linguistic macrofamily. Cambridge: The McDonald Institute for Archaeological Research.

Ruhlen, Merritt 1994. On the origin of languages Stanford: Stanford University Press.

Salmons, Joseph C., and Brian D. Joseph. 1998. Nostratic: sifting the evidence. Amsterdam and Philadelphia: John Benjamins.

Shevoroshkin, Vitalij V., and Thomas Markey 1986. Typology, relationship and time, Ann Arbor: Karoma.

Starostin, S.A. 1999. Subgrouping of Nostratic: comments on Aharon Dolgopolsky's The nostratic macrofamily and linguistic palaeontology In Renfrew and Nettle (eds.), pp 137–156.

Trask, R.L. 1996. The History of Basque. London. Routledge
———. 1999. Why Should a language have any relatives? In Renfrew and Nettle (eds.), pp. 157–76.

Watkins, Calvert. 1997. Thoughts on comparative linguistics in 1997. XVI International Congress of Linguists: Plenary Sessions, Paris.

ROBERT ORR

Louisiana Creole

Louisiana Creole (LC) is now a dying language, but remains spoken in the southern parts of the state. The number of native speakers of LC was estimated at somewhere between 60,000 and 80,000 in the early 1980s, but is no doubt lower today.

The first slave importations into what is today Louisiana took place in the second decade of the eighteenth century. The geographic origin of African slaves imported to French Louisiana is unusually well documented, and differs significantly from that of Caribbean slaves. In all, 64% were from Senegambia, 31% from the Bight of Benin, and 5% from Bantu-speaking areas, making Mande, West Atlantic, and Gbe languages the main substrates of LC.

It is not clear, however—and this has been subject to some debate—whether LC developed locally in Louisiana, or whether it was brought in from elsewhere, such as Africa or the Caribbean. On the one hand, the language does display similarities to West Indian French creoles, but no substantial immigration therefrom is attested until the beginning of the nineteenth century, when large numbers of refugees fleeing the revolutionary turmoil in what is today Haiti arrived. At this time, though, LC is generally believed to have already emerged.

Another migration of significance to the development of LC is the arrival from 1785 of groups of Acadians (ultimately) from Canada, who in Louisiana became known as *Cajuns*. The coexistence of LC as a minority language alongside Cajun French has had far-reaching consequences. Paradoxical though it may seem, given the obsolescence of Cajun French itself, most observers agree that LC decreolized and changed in the direction of Cajun during the twentieth century—many speakers have at the very least a passive understanding of Cajun French. Since the Cajun settlement has been limited along the Mississippi, this applies first and foremost to the prairie dialects rather than to the riverine varieties.

A final milestone, of course, in the history of LC is the transfer of the territory from French to American sovereignty in 1803. Thus began the process of complete replacement with English of French and LC as the majority language.

Although once more widespread, LC is today spoken in a couple of linguistic enclaves together constituting two main dialect areas—an eastern one along the Mississippi, and a western one around the Bayou Têche.

The best-studied riverine dialect is that of the upstream parish of Pointe Coupee, where there are perhaps 1,000 speakers of LC. Although there are speakers in New Orleans who have moved in from elsewhere in the state, the city has had an indigenous (rather Gallicized) dialect, the last few elderly speakers of which are still alive. Other varieties of the eastern

dialect are spoken in Vacherie and Kraemer (*Bayou-Bœuf*) in St. James. Another creolophone community relatively untouched by contact with Cajun or other varieties of French was found in the 1990s in St. Tammany just north of New Orleans. An offshoot of LC completely unaffected by Cajun has also been documented on the Mon Louis island in southern Alabama, but the last speaker of this variety died in the 1990s. Related varieties were spoken into the twentieth century in other localities in Mississippi, Alabama, and allegedly also in western Florida.

The nucleus of the FC-speaking area, however, is made up of the parishes of St. Martin, Lafayette, and Iberia in the Têche area. Children speaking or understanding LC may still be encountered in some communities in St. Martin, but they are becoming less and less numerous (they are no longer found in Pointe Coupee).

In addition, some few speakers are found in Lake Charles in the westernmost part of the state. Before World War II, LC could also be heard in the parishes of Natchitoches and Jefferson Davis. Outside of Louisiana, LC is also said to be spoken by emigrant communities Sacramento (California), southeastern Texas, and elsewhere.

Since most of the above-mentioned Haitian refugees and their slaves, just like the Cajuns, settled on the prairie rather than along the Mississippi, similarities with Haitian Creole are more salient in the riverine dialects than in the varieties centered around St. Martin.

It is believed that certain structural characteristics are due to the Haitian immigration, such as the possessive construction /zje a mwẽ/ 'my eyes' (alternating with presumably original /mo zje/) and the absolute possessive /mo kẽ/ 'mine' (cf. Northern Haitian /kẽ-a-m/).

As a result of the subsequent impact of Cajun, LC (even in its most basilectal forms) is—with the exception of Réunionnais—the French-lexicon Creole the least deviant from its lexifier. It is not obvious, though, that decreolization is the only factor responsible for this—it might be that a similar state of affairs was obtained even in the early days, and the structure of the Alabaman variety suggests that this is at least in part the case.

Some of the differences between the riverine Pointe Coupee variety and the more westerly Breaux Bridge (*Pont-Breaux*) dialect, which suggest a more far-reaching decreolization of the latter, include partial (re)introduction of grammatical gender, a more French-like system of adjectival comparison (using /ke/ instead of verb serialization involving /pase/ 'pass'), and changes in article usage and verbal morphology. Other comparatively recent changes in this direction include negation placement partly following French patterns, and a copula /(d)ɛt/ (< *d'être*). Some dialects also sport front rounded vowels and schwas, normally absent in French-lexicon creoles.

Other features that set LC apart from most of its relatives in the Caribbean include retention of a 2sg pronoun /to/, and preposed pronominal possessors. Contact with Cajun and English has, of course, left its mark on the LC lexicon as well, and Cajun is probably also responsible for the trilled realization of /r/ in LC.

In the communities of Vacherie and Kraemer, the white population has been heavily creolophone, although a shift to English is under way. A similar situation can be seen in villages in Pointe Coupee, St. John Baptist, St. Martin, Acadia, Lafayette, and St-Landry. In all, about a fifth of the LC-speaking population is believed to be made up of whites, but the white ethnolects have been shown to be even closer to French than their black counterparts.

References

Klingler, Tom. 1992. A descriptive study of the Creole speech of Pointe Coupee Parish, Louisiana. Ph.D. Thesis. University of Indiana.

Neumann, Ingrid. 1985. Le créole de Breaux Bridge, Louisiane. Hamburg: Helmut Buske Verlag.

Neumann-Holzschuh. 1987. Ingrid. Textes anciens en créole louisianais. Hamburg: Helmut Buske Verlag.

Valdman, Albert (ed.), 1997. French and Creole in Louisiana. New York: Plenum Press.

Valdman, Albert, Tom Klingler, Margaret Marshall, and Kevin Rottet. 1998. Dictionary of Louisiana Creole. Bloomington: Indiana University Press.

MIKAEL PARKVALL

Lyons, John

John Lyons is a specialist in semantics and linguistic theory. Since the publication of his *Introduction to theoretical linguistics* in 1968, Lyons has been recognized as an outstanding linguist. His major publications include *Introduction to theoretical linguistics* (1968), *Semantics* (2 volumes, 1977), *Language, meaning and context* (1980), and *Linguistic semantics* (1995). Other publications are *Chomsky*

(first published in 1970, with two other editions in 1977 and 1991) and *Language and linguistics* (1981). Lyons' work addresses a broad range of issues of fundamental importance to linguistics. Most of his books are characterized by a clear and accessible style that appeals to a wide student readership.

Introduction to theoretical linguistics covers all major aspects of theoretical linguistics, from phonetics, semantics, and syntax to structural linguistics, universal grammar, and generative and transformational grammar. Although the book presupposes no previous knowledge, it rigorously addresses a wide range of topics, bringing the reader to an advanced level of understanding theoretical linguistics.

John Lyons' volume on *Chomsky* (1970, 1977 2nd edition, 1991 3rd edition) is a guide to Noam Chomsky's life and ideas, particularly addressing his earlier groundbreaking work. Lyons emphasizes that the short and relatively nontechnical *Syntactic Structures* "revolutionized the scientific study of language".

John Lyons is particularly known for his seminal two-volume work, *Semantics* (1977). Volume 1 is a general and comprehensive introduction to the field of semantics. It brings together the results of studies on meaning and communication from many disciplines (including logic and philosophy), placing semantics in the larger context of semiotics. It deals with language as a semiotic system, behaviorist semantics, logical semantics, structural semantics, and semantic field theory. A broad review of current semantic theories is also found in *Language, meaning and context* (1981).

Volume 2 of *Semantics*, which can be read independently, deals with the main problematic issues in the semantics of natural languages: homonymy, polysemy, conversational implicature, deixis and anaphora, tense and aspect, mood and illocutionary force. The wide range of topics analyzed in this volume and the originality of the approach make the book a standard reference work in linguistics.

Designed as an elementary textbook, *Language and linguistics. An introduction* (1981) is a general introduction to linguistics and the study of language. It covers a broader range of topics than the classic *Introduction to theoretical linguistics* and presents the main subfields of linguistics: the sounds of language, grammar, semantics, language change, psycholinguistics, language, and culture.

Linguistic Semantics: An introduction (1995) is a critical discussion introducing linguistic semantics, defined by Lyons as the study of meaning systematically encoded in the vocabulary/grammar of natural language. The book was initially planned to be a second edition of his *Language, meaning and context* (1981). However, even though the general structure of the earlier book is preserved, this volume is very different from

its predecessor in that it covers several topics that were not previously discussed and takes into account new developments in the field. Its comprehensive style brings unequivocal definitions of specific, complex notions and explains the relationship between linguistic semantics and formal and philosophical semantics and contemporary pragmatics.

Linguistic semantics: an introduction begins with an explanation of what meaning is and introduces the standard distinctions between language and speech, *langue* and *parole*, competence and performance, form and meaning, and sentences and utterances. In addition, Lyons analyzes the different techniques that can be used to define the meaning of words. The book also deals with lexical semantics: it explains Peirce's distinction between tokens and types, as well as other fundamental notions such as homonymy, polysemy, synonymy, sense, reference, extension, intension, natural kinds, and prototypes. Lyons also discusses sentence meaning in terms of meaningful and meaningless sentences, implication and negation in natural language utterances, declarative interrogative, imperative, and exclamative sentences, and the principle of compositionality. It is the last part of the book, based on J.L. Austin's theory of speech acts, and particularly the chapter on context, that has been regarded as the most remarkable. Lyons believes that a satisfactory theory of context should include findings from social sciences such as psychology, anthropology, and sociology.

John Lyons' role in the history of linguistics has often been acknowledged. In 1995, a volume of essays by distinguished scholars was published to honor Lyons' outstanding contribution to linguistics. *Grammar and meaning*, edited with an introduction by F.R. Palmer, is a collection of essays related to the major issues studied by Lyons over the years.

Biography

Born in Manchester, UK, John Lyons studied at Cambridge, taught at London (1957–1961) and Cambridge (1961–1964), then became Professor of linguistics at Edinburgh (1964–1976) and Sussex (1976–1984), and Master of Trinity Hall, Cambridge (from 1984 to September 2000). In 1987, John Lyons was knighted for his exceptional contribution to the study of linguistics.

References

Lyons, John. 1968. Introduction to theoretical linguistics. Cambridge: Cambridge University Press.
———. 1977. Semantics, 2 vols. Cambridge: Cambridge University Press.

LYONS, JOHN

———. 1981. Language and linguistics. An introduction. Cambridge: Cambridge University Press.

———. 1981. Language, meaning and context. Glasgow: Fontana/Collins.

———. 1991. Chomsky, 3rd, revised and further enlarged edition. London: HarperCollins. First published, 1970.

———. (1995). Linguistic semantics: an introduction. Cambridge: Cambridge University Press.

Palmer, F.R. (ed.) 1995. Grammar and meaning. Essays in honour of Sir John Lyons. Cambridge: Cambridge University Press.

LAURA DANILIUC

M

Machine Translation

Exhilarating goals, serious disappointments, significant successes: all have accompanied machine translation in its twentieth- and twenty-first century existence. The early goal—fully automated translation of a quality equal to that achieved by human translators—has proven elusive, and some researchers doubt that it is feasible. Nonetheless, machine translation and its sister, machine-aided human translation, must be counted a success if measured by their effectiveness in extensive use by multinational companies, governments, and the European Commission. Computer tools that aid human translators by automating some parts of the translation process, coupled with machine translation programs whose output requires final human editing, have together met a worldwide demand that is increasing faster than the supply of human translators.

The history of machine translation in modern times begins in 1947 with a memo from Warren Weaver, then director of the Natural Sciences Division of the Rockefeller Foundation, to Norbert Wiener, a well-known computer scientist (then called cyberneticist): "I have wondered if it were unthinkable to design a computer which would translate. Even if it would translate only scientific material (where the semantic difficulties are very notably less), and even if it did produce an inelegant (but intelligible) result, it would seem to be worth while" (Hutchins 1997:197).

Only two years later, in 1949, the *New York Times* reported that the computer of the United States Bureau of Standards Laboratory, which it called the 'electric brain', would be able to translate foreign languages (Hutchins 1997:197–203).

Early machine translations proceeded word for word from the original language into the target language (in those days of the Cold War often Russian to English), a process that ignores the complexity of language structure and results in an inaccurate and barely readable translation. With the development of more sophisticated tools for computerized syntactic analysis, many researchers assumed that success was near at hand. Enthusiasm for the goal of automatic, high-quality machine translation fueled generous government and private funding of research and development of translation systems through the early 1960s. Dissenting voices such as that of Yehoshua Bar-Hillel (1960) that this perfectionist goal was unrealistic were at first not heeded.

In 1966, the National Academy of Sciences Automatic Language Processing Advisory Committee (ALPAC) reported that the prospects for fully automated high-quality, general-purpose machine translation were dim, causing immediate and massive cutbacks in funding for research in the field.

By the early 1970s, interest in the field revived, but accurate and publishable translation untouched by human hands was no longer the only goal; human–machine cooperation was envisioned and the creation of automated tools for translators became a special field within machine translation research.

The appearance of interactive desktop computer workstations greatly facilitated the development of translator's aids, and sophisticated 'translator's workbenches' were engineered. Current versions of these translator's workbenches include large databases of specialized vocabulary and high-quality desktop publishing tools, as well as 'translation memory', which enables the translator to store and reuse existing translations.

Currently used machine translation systems such as Logos, Metal, Systran, and PARS can translate even complex texts in a limited and defined domain, or subject area. Nonetheless, they require the services of humans to pre-or postedit the texts. In many important ways, the quality of machine translation has not risen appreciably since the 1970s, and the need for human editing is projected to continue.

At the other end of the spectrum, uses have been discovered for the rough, often inaccurate, and largely unreadable 'rough translations' that result when context is not specified and texts are not pre- or postedited. Such rough translations have proven economical in helping users to quickly categorize large numbers of foreign texts as either not useful or deserving of thorough translation.

Inexpensive translation software for personal computers, or freeware available on the Web, is widely used for obtaining the gist of e-mail, Web pages, and on-line discussion groups. These systems, often lacking large lexicons or phrase dictionaries, provide an even less exact gist than commercial systems. An example: French: *Il fait vraiment beau*! English: *He does indeed beauteous!* Correct translation : *The weather is really nice!* (Fortune City web site) But while many users are disappointed at the poor quality of these 'quick and dirty' translations, others are satisfied to have even an approximate idea of the contents of the text.

As always, the market is driven by demand. Currently, machine translation systems into and from English are available for a very large number of the world's languages. Translation systems into and from the major European languages (French, German, Italian, and Spanish) are available for a smaller number of world languages, while dramatically smaller numbers of systems are available to translate nonmajor languages into other nonmajor languages.

Spoken language, rather than written, is the focus of much current research. The Verbmobil project of the German Ministry of Research and Technology, for example, aims to prototype an on-the-spot oral translation of business negotiations between German and Japanese speakers. The problems of speech recognition and speech generation, however, are even greater than those involved in written translation. Such projects always aim at clearly defined topic areas and types of users, rather than at unrestricted use, but even with this limitation, it is not foreseen that large systems for oral translation will be available for several decades.

Other areas of current research include the development of desktop systems for the 'interactive' composition of specialized documents (for example, business letters) in a target language by a human who does not know that language. The system generates the text by a process of asking the user more and more detailed questions about what is meant, and choosing from its inventory of phrases and structure frameworks.

Machine translation continues to be a lively field for research and development. Although completely automated high-quality translation will probably not be achieved in the near future, more user-friendly, inexpensive systems will continue to be developed, particularly those aimed at special domains and for special purposes.

References

Arnold, Doug, et al. 1994. Machine translation. An introductory guide. Manchester/Oxford: Blackwell (url: http://clwww.essex.ac.uk/MTbook/).

Bar-Hillel, Yehoshua. 1960. The present status of automatic translation of languages. Advances in Computers 1. 91–163.

Booth, Andrew D. (ed.) 1967. Machine translation. Amsterdam: North-Holland.

Fortune City. 2001. Examples of nice translations made by automatic translators available on the market. http:// www.fortunecity.com/business/reception/19/mtex.htm

Henisz-Dosert, R., Ross Macdonald, and M. Zarechnak (eds.) 1979. Machine translation. The Hague: Mouton.

Hutchins, John. 1986. Machine translation: past, present, future. Chichester: Ellis Horwood.

———. 1997. From first conception to first demonstration: the nascent years of machine translation, 1947–1954. A chronology. Machine Translation 12. 195–252.

———. 2001. Home page. http://ourworld.compuserve.com/homepages/WJHutchins.

Kugler, M., K. Ahmad, and G. Thurmair (eds.) 1995. Translator's workbench: tools and terminology for translation and text processing. Berlin: Springer.

Machine Translation (journal), Dordrecht: Kluwer Academic Publishing, 1986.

Newton, J. 1992. Computers in translation: a practical appraisal, London: Routledge.

Whitelock, Peter, and Kieran Kilby. 1995. Linguistic and computational techniques in machine translation system design, 2nd edition. London: UCL Press.

RUTH H. SANDERS

Malagasy

Malagasy is one of two official languages of Madagascar (besides French). The population, estimated at 17.5 million for mid-2004, speaks either the standard language, a regional dialect, or both. Dialects also occur in Mayotte (Comoros).

History

Malagasy belongs to the Southeast Barito subgroup of the West Malayo-Polynesian subfamily of Austronesian. The language most closely related to Malagasy is Maanyan (in the southeast of Borneo).

Estimated dates of migration from Borneo range from 200 CE until the seventh century or later, based mainly on Sanskrit and Arabic loanwords exhibiting signs of borrowing via Malay. Non-Malay introductory passages in some seventh century Old Malay inscriptions have been tentatively identified as being in a precursory version of Malagasy, Maanyan, or close affiliate.

Bantu-speakers who subsequently arrived from the African mainland, adopting Malagasy, exercised substantial influence on its phonology and vocabulary. In the ninth to fifteenth centuries, contacts with Islamic traders led to considerable lexical borrowings from Swahili, Arabic, and Malay. In the fifteenth century, the Antaimoro in the southeast began writing, implementing the Arabic script soon also adopted by other dialects. In the early seventeenth century, the Menabe Sakalavas (western lowlands) gained hegemony over the west and northwest that lasted until the eighteenth century, promoting their dialect to corresponding prominence.

Around 1797, the reunited Merina in the highlands proceeded to expand, annexing the remaining highlands, the north and the east. Before the French conquest in 1895, the Merina kingdom encompassed the entire island except the south and part of the west. The Merina dialect (formerly Hova) was becoming standard in much of Madagascar, enhanced by schooling activities of British missionaries encouraged by Radama I (reigned 1810–1828). This also led to numerous English borrowings. The same missionaries introduced Latin-script spelling, made official by Radama I in 1820.

Under French rule (1896–1960), Merina remained the standard dialect of Malagasy, but the language of government, education, and the media was French. After independence, particularly after the "Malagasification" campaign launched in 1975, (Merina) Malagasy became the language of school education, and the predominant language of government and the media. But French remains the language of higher education.

This century-long bilinguality led to profuse borrowing from French.

Dialects

Malagasy has a tripartite division into East (geographical north, center, east, and southeast), West (west and southwest), and Intermediate (Betsileo, Bara, and Antanosy, sharing features of both former) groups. The following table compares Menabe Sakalava (West dialect, henceforth 'Sakalava') with Merina (East) and the reconstructed ancestor language Proto-Malagasy (P-Mlg):

P-Mlg > West/East	P-Mlg	Sakalava	Merina	English
*li- > li/di	*lime	lìme	dìmy	'five'
	*kulit	hòlitse	hòditra	'skin'
*ele > ele/eli	*kelek	hèleke	hèlika	'armpit'
*ti- > ti/tsi	*tina'i	tinày	tsinày	'intestines'
	*witik	vìtike	vìtsika	'ant'
*te > te/ti	*katen	hàte	hàtina	'itch'
-m,-n,*ŋ > -Ø/-na	*lalan	làla	làlana	'path'
*-r, *-t > -tse/-tra	*epat	èfatse	èfatra	'four'

Phonology

Merina Malagasy has 19 primary consonantal segments, spelled p, t, k, b, d, g, f, s, h, v, z, l, r, m, n, ts, j /dz/, tr /tɻ/, and dr /dɻ/ (including four affricates), and nine prenasalized consonants functioning as single segments: mp, nt, nk, mb, nd, ng /ŋg/, nts, nj, ntr, ndr. Other dialects additionally have velar ñ /ŋ/. In southern central-highlands dialects, ts is /ð/. In Betsimasaraka, s is /ʃ/. Some base-initial consonants undergo nasal sandhi with the final nasal of prefixes.

Bisegmental consonant clusters are mainly limited to loanwords and include r, l, or s as one component (artìsta 'artist', jeografia 'geography', anglìsy 'English'), also word-initially (frantsày 'French', krismàsy 'christmas'). Frequently, clusters were separated by an anaptyctic i (cf. adirèsy 'address', alikàola 'alcohol', Madagasikàra 'Madagascar') or, by vowel harmony, an o (boròsy 'brush', dokotèra 'doctor') or a (kàratra 'card'). But in casual speech, an unstressed interconsonantal vowel is sometimes dropped (fòtsiny [fùtsini] > [fùzni] 'only').

There are five simple vowels spelled *a*, *e*, *i* or *y*, *o* /u/, and *ô* /o/, and six diphthongs, *ai* or *ay* /aʸ/, *ao* /aʷ/, *ià* /iˀ/, *ià* /ʸa/, *òa* /uˀ/, and *oà* /ʷa/ (*y* stands for *i* word-finally). The *ô*, restricted to loanwords (*hôtèly* 'hotel', *tômôbìlina* 'automobile'), is often printed without diacritic, leading to ambiguity. In casual speech, the dipthongs can contract to [e] (< *ai, ià*), [i] (< *ìa*), [o] (< *ao, oà*), or [u] (< *òa*).

Malagasy does not normally have final consonants. Historical word-final coda were either dropped (*lepas > lèfa* 'liberation, escape'; *lalow > làlo* 'pass'; *apuy > àfo* 'fire') or retained as 'weak ending' with a dummy vowel spelled *a* (in some dialects *e*), pronounced as short schwa [ă] or [ɜ]: *-k, *-p > *-ka*; *-r, *-t > *-tra*; *-m, *-n, *-ŋ > *-na* (some dialects > *-ña,* and some deleted it). At suffixation, the historical situation is variedly reflected in the morphophonemics (*lèfa → fandefàsana* 'broadcast'; *làlo → lalòvina* 'passed'; *tsìdika → fitsidìhana* 'short visit'; *lèlaka → lelàfina* 'licked').

Stress (mainly indicated in dictionaries and language primers) is distinctive, cf. *tànana* 'hand' (< *taŋan) vs. *tanàna* 'town' (< *tana'-an). Words with a 'weak ending' have either stress on the antepenultimate syllable that shifts upon suffixation (*àkatra → akàrina* 'be raised') or penultimate stress that does not (*ja'it > zàitra → zàirina* 'be sewn'; *ta'an > tàna > fitànana* 'endurance'). Words without a 'weak ending' have either penultimate stress that shifts upon suffixation (*fèno → fifenòana* 'filling'; *dìa → fandiàvana* 'threshold') or ultimate stress that does not (*umei > omè → fanomèzana* 'gift'), unless an additional vowel appears (*behay > bè → fanabeàzana* 'upbringing'). When the stress shifts onto an *i/y*, this becomes *e* either when the preceding vowel is *e* (*kèly → fikelèzana* 'reduction'), or sometimes reflecting historical *e (*alem > àlina → fialèmana* 'dinner time').

Morphology

The morphology of Malagasy involves flexion (e.g. in tense alternation), affixation (voice markation, nominal derivation), alternating clitics (personal pronouns), and root alternation (deictics).

The noun is practically invariant. Number is indicated with a complex system of demonstratives that also expresses proximity: *ìo pàoma ìo* 'this apple'; *irèo pàoma irèo* 'these apples'; *irèry pàoma irèry* 'those apples'; *Marikìvy ìo* 'This one is sour.' There is also a series of locatives for 'visible' with *i-* (*ìo, ìty, ìto, ìtsy, iròa,* etc.), somewhat visible with *e-* (*èo, èty, ...*), or vague with *a-* (*ào, àty, ...*), etc. The definite articles are *ny* for nouns, and *i* for proper names.

Personal pronouns distinguish nominative (subject), oblique (object, possessive pronoun), and genitive forms (possessive adjective, actor-argument

of root-passive verbs); the latter attaches to a host word similar to a suffix:

Person and Number	Nominative	Oblique	Genitive
1st singular	*(iz)àho*	*àhy*	=*ko*/=*o*
2nd singular	*ianào*	*anào*	=*nào*/=*ào*
3rd singular	*ìzy*	*àzy*	=*ny*/=*ny*
1st plural inclusive	*isìka*	*antsìka*	=*ntsìka*/=*tsìka*
1st plural exclusive	*izahày*	*anày*	=*này*/=*ày*
2nd plural	*ianarèo*	*anarèo*	=*narèo*/=*arèo*
3rd plural	*ìzy irèo*	*àzy irèo*	=*n'ìzy irèo*/=*ny*

Authentic adjectives (*kèly* 'small', *màina* 'dry', *tsàra* 'good') are distinguished from adjectival verbs: the latter feature tense (*madìo* 'is clean', *nadìo* 'was clean', *hadìo* 'will be clean') and voice (*diòvina* 'be cleaned'). Both form degrees of comparison via combination with independent markers: *tsàra kokòa* (good more) 'better', *tsàra indrìndra* (good extremely) 'best'. In the comparative, the marker is optional: *Natsìro [kokòa] nòho ny poàra ny pàoma* (was-tasty [more] than the pear the apple) 'The apple was tastier than the pear.'

The numerals are *irày, ròa, tèlo, èfatra, dìmy, ènina, fìto, vàlo, sìvy* '1...9', *fòlo* '10', *ròapòlo* '20', *tèlopòlo* '30', etc., *zàto* '100', *ròan-jàto* '200', etc. Teens are formed by adding *ambìn'ny fòlo* after the digit (but *iràika ambìn'ny fòlo* '11'). Numerals follow numerated nouns (*òlona tèlo* 'three **persons**'), but precede units of measure (*pàoma tèlo **kilào*** 'three **kilos** of apples').

The verbs have an elaborate paradigm of forms that does not include a plural.

Tense is expressed by alternating an optional prefixal *m-* (present) with *n(o)-* (past) or *h(o)-* (future): *manadìo* 'cleanse'/*nanadìo* 'cleansed'/*hanadìo* 'will cleanse'; *mitèny* 'speak'/*nitèny*/*hitèny*; and *indràmina* 'be borrowed'/*nindràmina*/*hindràmina*. The optional *o* appears when the prefixless present-tense form has an initial consonant: *soràtana* 'be written'/ *nosoràtana*/ *hosoràtana*.

Voice is intertwined with some other categories. The active voice is formed with the prefix *maN-* (action verbs), *mi-* (action and status verbs), or *ma-* (status and adjectival verbs) to the stem (***manòratra** àho* 'I **write**'; ***Mitèhina** ny làkana àho* 'I **shove** the boat [with a pole]'; ***Matòry** àho* 'I **sleep**'; ***Mànana** tràno aho* 'I **have** a house'). Causative verbs prefix *maha-* (*fàly* 'happy' → *mifàly* 'rejoice, be merry' → *mahafàly* 'be pleasing'; *tònga* 'reach' → *mahatònga* 'cause'). The active imperative has the suffix *-a* (*manoràta* 'write!'; *mitehèna* 'shove! [with pole]'; *matorìa* 'sleep!').

Passive forms are frequently used and the most variegated. There are some 30 so-called root-passive

status verbs that topicalize the undergoer in an uninflected form (*àzo* 'received, understood'; *hèno* 'listened to', *hìta* 'seen'). A pronominal actor is in the genitive: *Hènoko ìzy* 'He is listened to by **me**.'

Passive forms of prefixed-active verbs have the suffix *-ina* or *-ana*: *Tehènina ny làkana* 'the boat **is shoved** [with a pole]'; *Totòfana tàny ny làvaka* 'the hole **is filled up** with earth'. Pronominal actors are in the genitive: *Soràtano ny taratàsy* 'The letter is written by **me**.' Active derivations of root-passive verbs have suffixed passives (*tsìnjo* 'be perceived/ noticed' → *mitsìnjo* 'observe/ watch' → *tsinjòvina* 'be observed/ watched').

For verbs with two nonactor arguments, a 'stationary' (medium, goal) and a 'mobile' one (instrument, theme), the former can be topicalized either by the above passives, or with the so-called circumstantial passive, formed by circumfixation of *aN-...-ana* or *i-...-ana* (*mandràkotra* 'cover' → *andrakòfana* 'be covered [with]'; *vìdy* 'buy' → *ividìana* 'have [something] bought [for one]'). The 'mobile' argument is then topicalized by a passive with prefix *a-* (*mandròso* 'serve [food] [to someone]' → *rosòana* 'be served [guest]'/*aròso* 'is served [food]').

The above passives all partake in the (*m*)/*n*(*o*)/h(*o*) tense alternation and form imperatives with *-o* (or *-y* when the base has *-o-*): *henòy* 'listen!'; *andràsana* 'be waited for' → *andràso* 'wait!'; *rosòy* 'serve [to someone]!'/*arosòy* 'serve [a dish]!'.

Finally, there is the resultative passive with *voa-* (*mandàza* 'say, declare' — *voalàza* 'be/get said'; *mifìdy* 'choose' — *voafìdy* 'get chosen') and the incidental resultative with *tafa-* (*mivèrina* 'return' — *tafavèrina* '[unexpectedly] be/get back'). Both have a future (with *ho-*) but no past tense, and have no imperative.

Nominalization of verbs is productive. Agent nouns derive from active verbs by inserting *-p-* after the prefixal *m-* (*mandèha* 'go' → *mpandèha* 'passenger, voyager'; *mitèny* 'speak' → *mpitèny* 'speaker'). Nouns denoting some aspect of that which is expressed by the verb can be formed by replacing prefixal *m-* by *f-* (*fandèha* 'way of walking'; *fitèny* 'manner of speaking'). Abstract nouns are formed by additional suffixation of *-(an)ana* (*fambolèna* 'cultivation'; *fidìrana* 'admission'; *fandehànana* 'departure', *fitenènana* 'conversation').

Syntax

Typically, an attribute follows the modified element—*ny tràno lehibè* 'the **big** house'—while a predicate precedes the subject—*Lehibè ny tràno* 'The house is **big**'—also demonstrating that existential clauses are formed without a 'to be' equivalent.

In possessive atribution with nominals (pronouns see above), the genitive marker *n′* is preposed to the definite article (*n′ny bòky* 'of the book', *n′i Rìta* 'Rita′s').

Names with the article *Ra-* undergo nasal sandhi (*n+r > ndr*), and a hyphen replaces the apostrophe (*ny bòkin-dRasòa* 'Rasoa's book'). Preceding adjectives merge with the genitive marker: *ny bòky vaovàon′ny mpampiànatra* 'the teacher′s **new** book'.

Malagasy sentences are mainly predicate-initial: *Vaovào ny bòky* 'The book is **new**'; *Àhy ny bòky vaovào* 'The new book is **mine**'; *Mpampiànatra àho* 'I am a **teacher**'; *Tèlo ihàny ny bòky* 'There are only **three** books'; and *Mitèny àho* 'I **speak.**' Transitive sentences are typically Verb–Object–Subject (VOS): *Mivìdy ny bòky àho* '*I buy the book*' and *Manòratra taratàsy i Rìta* 'Rita **writes** a letter.' Indirect objects precede the subject: *Mivìdy bòky ho an′ny zàza i Rìta* 'Rita buy**s** a book **for the child**.'

Subject fronting with emphasizing is achieved with the particle *no* (*Ny mpampiànatra no nivìdy ny bòky* '**The teacher** [is the one who] bought the book') that emphasizes any fronted term: *Omàly no novidìna ny bòky* 'The book was bought [precisely] **yesterday**.' Subject fronting with the particle *dìa* emphasizes the verb: *Ny mpampiànatra dìa nivìdy ny bòky* 'The teacher **bought** [didn′t steal] the book.' It also emphasizes a sentence-final subject: *Hivìdy ny bòky dìa ny mpampiànatra* '**The teacher** [noone other] will buy the book.'

The relative marker is *izày*: *ny bòky izày novidìn′i Rìta* 'the book **that** was bought by Rita.' It is often optional: *Hanòratra taratàsy ny mpampiànatra [izày] nivìdy ny bòky* 'The teacher **who** bought the book will write a letter.'

The above article still leaves out many features, particularly numerous irregularities and exceptions, but also various intricate regularities, that make Malagasy an extremely interesting object of linguistic study.

References

Adelaar, Alexander K. 1995. Asian roots of the Malagasy: a linguistic perspective. Bijdragen tot de Taal-, Land- en Volkenkunde 151. 325–56.

Andriamanantsilavo, Seth, and William Ratrema. 1981. Ny fitsipi-pitenenantsika (The grammar of our language), Vol. I. Antananarivo: Librairie mixte.

Dahl, Otto C. 1991. Migration from Kalimantan to Madagascar. Oslo: Norwegian University Press.

Mahdi, Waruno. 1988. Morphophonologische Besonderheiten und historische Phonologie des Malagasy (Morphophonological particularities and historical phonology of Malagasy). Berlin and Hamburg: Reimer Verlag.

Paul, Ileana (ed.) 1998. The Structure of Malagasy, vol. II, Los Angeles: UCLA Department of Linguistics.

Pearson, Matthew, and Ileana Paul (eds.) 1998. The structure of Malagasy, Vol. I. Los Angeles: UCLA Department of Linguistics.

Rajaonarimanana, Narivelo. 1994. Grammaire moderne de la langue malgache. Paris: L'Asiathèque.

Randriamasimanana, Charles. 1999. Clausal architecture and movement verbs in Malagasy. Selected papers from the Eighth

International Conference on Austronesian Linguistics, ed. by E. Zeitoun and P.J.-K. Li. Taipei: Academia Sinica.

Rasoloson, Janie, and Carl Rubino. 2004. Malagasy. The Austronesian languages of Asia and Madagascar, ed. by A.K. Adelaar and N. Himmelmann. London and New York: Routledge.

Schmidt, Bernd. 1984. Die Malgassisierung und das 'malagasy iombonana'— Bilanz der Verwendung und Schaffung einer Nationalsprache (Malagasification and the 'communalization of Malagasy' campaign — net results of implementation and creation of a national language). Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikations for schung 37. 618–29.

Simon, Pierre. 1988. Ny fiteny fahizany, reconstitution et périodisation du malgache ancien jusqu'au XIVe siècle. Paris: Institut des Langues et Civilisations Orientales.

Vérin, Pierre, Conrad P. Kottak, and Peter Gorlin. 1969. The glottochronology of Malagasy speech communities. Oceanic Linguitics 8. 26–83.

<div align="right">WARUNO MAHDI</div>

Malay-Indonesian and Malayic languages

Malayic is a subgroup of the Western Malayo-Polynesian branch of the Austronesian language family. Malayic languages are spoken throughout the Malay-Indonesian archipelago, from Sumatra in the west to New Guinea in the east, and also on mainland Southeast Asia, mostly in the Malay Peninsula and in parts of Thailand.

Malay-Indonesian is a member of the Malayic subgroup. The indigenous name for the language is *Bahasa Melayu* (literally, 'the Malay language'), but some varieties used in Indonesia are also called *Bahasa Indonesia* ('the Indonesian language', see below). Similar forms of Standard Malay-Indonesian serve as the national languages of Indonesia, Malaysia, Brunei, and Singapore; the latter three are particularly similar to each other. A large number of Malay dialects and Malay-based pidgins and creoles are also spoken throughout the region.

The number of speakers of Malay-Indonesian is estimated at over 250 million, making it by far the most widely spoken language in Southeast Asia, and a major world language. Most speakers do not use Malay-Indonesian as a native home language; the number of native speakers is difficult to estimate, but is probably over 50 million. This figure is rapidly increasing, as more and more people in Indonesia, Malaysia, and Brunei shift from their ancestral home languages to Malay-Indonesian.

Classification

While there is wide agreement about the existence of the Malayic subgroup and the Malay-Indonesian language, linguists have not been able to agree on their classification. Malayic used to be classified together with Javanese, Sundanese, Madurese, Achenese, and Lampung as in a putative 'Malayo-Javanic' branch, but strong doubts have been cast on the validity of this classification. It is now clear that Malayic is more closely related to the Chamic languages, spoken in Cambodia and Vietnam, than to any of these languages. One factor hindering classification is that many languages have borrowed heavily from Malay-Indonesian, which has served as a regional lingua franca for many centuries. This may make such languages appear to be much more closely related to Malayic than they actually are. Similarly, no clear criteria have been established for distinguishing between Malayic languages and dialects of Malay-Indonesian. Scholars in the field have therefore preferred using the neutral term 'isolect' to refer to any Malayic speech form which has a name of its own and is regarded by speakers as distinct from other varieties.

There is great variation among Malayic isolects, many of which are not mutually intelligible. They fall into several categories. Some, like Riau Malay (spoken in the Riau archipelago in Indonesia) or Kedah Malay (spoken in the Malaysian state of Kedah), are thought to be direct descendants of Proto-Malayic, a hypothetical language reconstructed on the basis of modern isolects. Other isolects, however, have had a more complex history, and owe their emergence to language contact and language shift. For example, Betawi, the language of the indigenous ethnic group of Jakarta, is based on Malay, but has incorporated lexical and grammatical elements from Balinese, Javanese, Sudanese, and Chinese, languages spoken by the ancestors of today's speakers. Some isolects have developed from pidginized forms of Malay, collectively known as Bazaar Malay, which originally served only for interethnic communication, not as a first language. Baba Malay, spoken by acculturated Chinese communities in Malacca, Penang, and Singapore, is thought to have developed from Bazaar Malay, which gradually became the speakers' first language. Most isolects spoken in eastern Indonesia are also believed to have developed from forms of

Bazaar Malay. This complex situation has also contributed to the difficulty in classifying Malayic isolects.

Standard Malay-Indonesian

History

The cradle of Malay civilization and of the Malay-Indonesian language was in south-central Sumatra. Many scholars believe, however, that its hypothetical ancestor Proto-Malayic was spoken in western Borneo. The original place name Malayu (=Malay) has been identified with the former Malay kingdom of Jambi in central Sumatra. The Chinese monk I Ching, who visited the area in the seventh century AD, reported on a place called 'Moloyu'; later Javanese inscriptions and manuscripts also refer to the area of Jambi as 'Malayu'. The *Sejarah Melayu* ('Malay Annals'), the canonical work of classical Malay literature, traces Malay origins to Palembang, a city south of Jambi, which historians and archeologists identify with the center of the ancient maritime empire of Sri Vijaya.

The earliest direct evidence of Malay comes from a handful of seventh-century AD inscriptions, found in southern Sumatra and on the nearby island of Bangka, and associated with Sri Vijaya. While not all scholars agree that the language of these inscriptions is the direct ancestor of modern standard Malay-Indonesian, it is commonly referred to as Old Malay. The inscriptions are written in a formal language that borrows heavily from Sanskrit; there is no direct evidence of the language ordinary people spoke in their everyday lives. Old Malay inscriptions dating from the ninth century were also found in areas where Malay was not indigenous, like Java and the Philippines, showing the early spread of Malay as a literary language throughout the region.

The use of Old Malay as a literary language in Java and in the Philippines did not survive long. However, in Sumatra it has continued uninterrupted. Even among peoples who speak rather different languages, like Rejang and Lampung, Malay (written in an Indian-derived script called *rencong*) continued to serve as the major literary language. The oldest extant Malay manuscript is a recently rediscovered fourteenth-century work originating from southern Sumatra. Some letters and longer works from the sixteenth century are preserved in collections in the West, and from the seventeeth century onward, Malay manuscripts become numerous. The contents of these works are varied, and range from legends, chronicles, and religious treatises, to legal documents and letters. The language of these manuscripts, while showing some variation across time, space, and style, is nevertheless remarkably uniform, and has been termed Classical Malay.

Modern Standard Malay-Indonesian came into existence in the nineteenth century, as a joint (although

sometimes uncoordinated) effort of native and Western scholars. The great Malay scholar Raja Ali Haji (c. 1809–1870) composed a grammar and a dictionary of Malay. Later, the Dutch scholar C.A. van Ophuysen (1854–1917) standardized the language for use at schools throughout the Dutch Indies. In 1928, a congress of nationalist students declared the Malay language, under the name *Bahasa Indonesia*, as the national language of the Indonesian nation. During the Japanese occupation (1942–1945), the modernization of Malay-Indonesian received an impetus, as it was widely used in the administrative and educational systems and in the mass media. Indonesia declared its independence in 1945, whereby Malay-Indonesian (under the name *Bahasa Indonesia*) became its national language. When Malaysia, Singapore, and Brunei followed suit, Malay-Indonesian (under its traditional name *Bahasa Melayu*) became their national language as well. In 1972, the spelling of Malay-Indonesian was reformed and harmonized, and a joint council (known by its acronym MABBIM) has been coordinating language planning activities in these countries ever since.

Typical Features

The sound system The sound system of Standard Malay-Indonesian is relatively simple. There are 18 primary consonants (given in their conventional orthography): *b, d, j, g, p, t, k, c, m, n, ny, ng, l, r, h, s, w, y*. Sometimes the glottal stop is also analyzed as a phoneme. (In this orthography, *c* and *j* represent palatal stops, while the digraphs *ny* and *ng* represent a single consonant each, the palatal nasal and velar nasal, respectively). The vowels are *a, e, i, o, u,* plus a mid-central vowel (confusingly written *e,* like the mid-front vowel). There are also a few loan phonemes, like *f* and *z*, which first entered the language via loanwords. In Standard Malay-Indonesian, there is no distinctive word accent.

Word formation Beyond borrowing and coining new roots, Malay-Indonesian words are formed by three principal processes: affixation, reduplication, and compounding. Affixes include prefixes (*makan* 'eat' vs. *di-makan* 'eaten'), suffixes (*kerja* 'work' vs. *kerja-kan* 'do [something]'), and circumfixes, which attach themselves simultaneously to the beginning and end of roots (*baik* 'good', *ke-baik-an* 'goodness'). Redupli-cation fills many functions. For example, *anak* means 'child', while *anak-anak* means 'a group of children'; *jalan* means 'walk' or 'road', *jalan-jalan* is 'go for a walk, go out'. Compounding combines two existing roots into one word; for example, *matahari* 'sun' is made up of the words *mata* 'eye' and *hari* 'day', and literally means 'the eye of the day'. In addition to these three processes, clipped words (made up of parts of existing words) and acronyms are very common, especially in Indonesia.

Word classes Most scholars agree that Malay-Indonesian has both function words and content words, and that content words include at least two categories: nouns and verbs. Some analyses also posit adjectives, adverbs, and other classes. Part of the discrepancy stems from different semantic, morphological, and syntactic criteria used by different scholars to classify words. Another factor has been Eurocentricity: scholars (especially in the past) tried to fit Malay-Indonesian into categories created by European scholars for European languages, which were often incompatible with the actual categories of Malay-Indonesian. In fact, it is possible that Malay-Indonesian roots are precategorial, with the category or word class determined by affixes or by the place the word fills in the sentence.

Word order The basic word order in Malay-Indonesian is, like English, Subject–Verb–Object: *Tuti makan pisang* (lit. *Tuti eat banana*) 'Tuti is eating bananas.' However, according to the circumstances and the particular emphasis desired by the speaker, these constituents can be ordered in any way. Also, unlike in English, the subject and the object can be left unsaid, if they are already understood (or deemed so by the speaker). Focusing on the object (as in English passive sentences) and not on the subject is often the normal way of expressing things in Malay-Indonesian. Thus, the usual way to say that 'someone ate the bananas' would be *pisang itu dimakan (orang)*, which literally means something like 'the bananas were eaten (by people)'. Modifiers normally follow the head, as in *rumah besar* (lit. *house big*) 'a big house' and *rumah itu* (lit. *house that*) 'that house'. Possession can be indicated by a simple juxtaposition of nouns: *rumah Tuti* (lit. *house Tuti*) 'Tuti's house', *rumah saya* (lit. *house I*) 'my house'.

An important difference between Malay-Indonesian and Indo-European languages like English is that words are not inflected, and many categories expressed by inflection in Indo-European languages are simply left unspecified in Malay-Indonesian. Thus, *rumah* can mean either 'house' or 'houses', and *makan* can mean 'eat', 'eats', 'am eating', 'have eaten', 'ate', etc., depending on the context.

Writing Systems

All early scripts that were used for writing Malay were based on Indian writing systems. The oldest Malay inscriptions were written in a slightly modified version of the Pallava script of southern India. The earliest Malay manuscript was written in the Kawi script, used for writing Old Javanese, while later Malay manuscripts from Sumatra were written in the local *rencong* script, an angular alphabet also ultimately derived from Indian writing systems.

During the first half of the second millennium AD, the Malays, who were originally animists and then Hindus and Buddhists, gradually converted to Islam. With the new religion, they also acquired the Arabic alphabet, which formed the basis for the Malay alphabet known as Jawi. The earliest example of Jawi writing is a fourteenth-century inscription from Terengganu in the Malay peninsula, and the earliest Jawi manuscripts are two letters written by the sultan of Ternate to the king of Portugal in 1520 and 1521. Since the seventeenth century, Romanized alphabets have gradually replaced Jawi, but Jawi is still used sometimes, especially in Malaysia and Brunei.

Other Malayic isolects

The number of distinct Malayic isolects probably runs into the hundreds. A few of the major ones are listed below.

Iban: The Iban people of northern Borneo were formerly known as Sea Dayaks. They speak a number of closely related isolects, mostly in the Malaysian state of Sarawak, but also in the neighboring state of Sabah and across the border in Kalimantan (Indonesian Borneo). The total number of speakers is estimated at upward of 500,000. Iban is more divergent from Malay-Indonesian than other major Malayic isolects. For this reason, it is often classified as a separate language.

Minangkabau: This important Malayic isolect is used by about 7 million people. Native to western Sumatra, it is now spoken via immigration outside its historical homeland, principally in the Indonesian province of Riau and in the Malaysian state of Negri Sembilan. Often classified as a separate language, Minangkabau has a long literary tradition, and has influenced standard Indonesian through the numerous authors of Minangkabau origin who played an important role in the creation of modern Indonesian literature.

Banjarese: Banjarese probably came into existence as a lingua franca used among the Dayaks and Malays of southern Borneo and their Javanese overlords. It is now spoken as a first language by several million people in the city of Banjarmasin and in other locations throughout the southern parts of Borneo.

Riau-Johor Malay: This dialect group is spoken in the southern part of the Malay Peninsula and in the Riau archipelago in Indonesia. Of all modern vernaculars, it bears the closest resemblance to the national standard language of Malaysia, Singapore, and Brunei.

Kelantan-Patani Malay: This group of closely related dialects is spoken by several million speakers on both sides of the Thai–Malaysian border. Kelantan-Patani Malay is characterized by radical phonological changes, which make it difficult or impossible to understand by speakers of other Malay dialects.

Ambonese Malay: This creole is spoken natively on the island of Ambon in Maluku, where it is in the process of replacing the indigenous languages. It is also widely used as a lingua franca (and increasingly as a first language) in other parts of Maluku. Ambonese Malay probably developed from a form of Bazaar Malay, which eventually became the first language of some speech communities. Closely related isolects are spoken in and around Kupang (on the island of Timor) and Larantuka (on the island of Flores).

Manado Malay: This creole originated on the island Ternate, as evidenced by its Ternate lexical component. It is spoken natively in the city of Manado in North Sulawesi, and is gradually replacing the local Minahasa languages throughout this province. Closely related varieties are spoken in northern Maluku and in many parts of northern and central Sulawesi.

Jakarta Indonesian: The urban koine of Jakarta is based on Betawi, the Malayic isolect used by the indigenous inhabitants of Jakarta, but has been influenced by Standard Indonesian. With at least 5 million native speakers and many more second-language speakers, it is one of the largest Malayic isolects. Jakarta Indonesian is becoming the language of choice for educated urban speakers throughout Indonesia.

References

Adelaar, K. Alexander. 1992. Proto-Malayic: the reconstruction of its phonology and parts of its lexicon and morphology. Pacific Linguistics Series C-119. Canberra: The Australian National University.

Asmah Haji Omar. 1983. The Malay peoples of Malaysia and their languages. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Collins, James T. 1998. Malay, world language: a short history. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Sneddon, James Neil. 1996. Indonesian reference grammar. St. Leonards, NSW: Allen & Unwin.

URI TADMOR

Malkiel, Yakov

Born in Ukraine, Yakov Malkiel was formed intellectually in Germany, where he received his doctorate in Romance Linguistics. With the beginning of World War II, he moved to the United States, the country in which he combined his work as a Professor of Romance Philology in the University of Berkeley (California) with the development of most of his scientific career.

His wide cultural and linguistic knowledge, his enormous capacity of work, and his astonishing domain of several languages allowed him to produce an enormous bibliography, written in German, Italian, French, Portuguese, Spanish, and English. Although this bibliography only includes a dozen books, it contains several hundred articles and essays. In fact, Malkiel himself published his own *Autobibliography* in 1988 (*Yakov Malkiel: a tentative autobibliography*), in which more than 800 titles are registered. Such a vast scientific production comprises, as expected, very diverse topics, which makes it difficult to find a superficial description for this linguist's work lines and contributions. Nevertheless, as a guideline, it can be affirmed that the scientific work of Yakov Malkiel is related to three large thematic fields that, presented from less to more importance, are:

In the first place, it is related to the historiography of Linguistics, the field in which he published essays on the appropriate methodology of this discipline, on some of the schools and tendencies of modern Linguistics, and on some of their main figures (Leonard Bloomfield, Leo Spitzer, etc.).

It is also related to the general synchronic Linguistics, the study of (a) language(s) at a given point in time. In this sense, it is necessary to highlight his interesting studies on what he denominated *irreversible binomials*, that is, those idiomatic expressions formed by two elements whose order cannot be permuted (*odds and ends*, but **ends and odds*); on the special uses that the languages make of the letters (in the construction of acronyms, in the construction of fixed expressions, etc.); or on the relationship that exists between grammatical gender on the one hand and sex and size on the other.

But the scientific effort of Yakov Malkiel was primarily centered on diachronic (historical) linguistics, the field in which, in turn, we may distinguish two areas:

In the first place are his works on the methodology and epistemology of diachronic Linguistics and its different fields of study (etymology, forms and causes of the linguistic change, studies on phonology and historical morphology, etc.). In this area, Malkiel tried to give a new impulse to a type of linguistic investigation that was in clear decadence after the almost absolute triumph, at the beginning of the twentieth century, of the synchronic perspective defended by

Ferdinand de Saussure and his followers (the so-called structuralists) and, later, by the Generativism spawned by Noam Chomsky.

To get that revitalization of the historical studies, he established new conceptual distinctions, separating the true diachronic Linguistic, that according to him should focus on the study of specific linguistic processes, from the *Glottodynamics*, that should study the universal constants that rule linguistic change and evolution. Moreover, he substituted the proper term "diachronic Linguistics" or "historical" for that of *genetic Linguistics*, convinced that evolutionary Linguistics should not be limited to describe mere historical developments, but rather that it has to be able to discover the decisive causes of these developments. Along those lines, his theory on the *multiple causation* of linguistic change is especially remarkable, according to which most of these changes have their origin in several interrelated factors, factors that can depend on different linguistic planes (phonology, morphology, grammar, or semantics) or, even, in the extralinguistic reality.

At the same time, as he himself affirms in the foreword of one of his most important works (*From particular to general Linguistics*), he tried to harmonize genetic Linguistics with the new theoretical perspectives coming from structuralism, thus changing the atomistic type of study developed by traditional historical Linguistics for another that is more in accordance with the general lines of contemporary linguistic thought. In fact, it could be affirmed that Malkiel was, because of his way of analyzing the data and of understanding the reality of language, among the structuralist followers of Bloomfield, although he never lost his scientific individuality and maintained, with respect to the latter, important theoretical differences.

In the second place, the dedication of Malkiel to the field of historical Linguistics is reflected in the publication of countless specific studies on diachronic phonology and morphology. In those works, almost always centered on the Roman languages—and, among those, especially Spanish, Portuguese, and Italian—he tried to combine the explanation of particular facts with the formulation of general theories, something that connects these works with those described previously and that also allows his results to be extrapolated to the analysis of any other language or group of languages. From among those essays, one can cite the ones dedicated to the explanation of the evolution of certain sounds or groups of sounds in the Romance languages: those related with symbolic values of the sounds (phonosymbolism) and, above all, those focused on the study of diverse elements derivative from the Romance languages. Among the latter, those in which he argues about the possible existence of diverse types of empty morphs (morphologic elements lacking of meaning) in the Romance languages are of special importance, like the case of those that he denominated *hispanic interfixes*: morphemes without meaning that would appear between the root and the suffix of certain Spanish words (like *pan-ad-ero* 'baker', where *pan* is the root that means 'bread', *-ero* is a suffix that expresses 'agent', and *-ad-* is an intermediate element lacking any meaning).

In definitive, Yakov Malkiel can be characterized as one of the main figures of contemporary diachronic Linguistics, to which he dedicated most of his efforts and his vast knowledge, something noticeable in his numerous publications which are so filled with knowledge and data that they are quite inaccessible to the uninitiated in the particular subject matter. This dedication to a linguistic area that has been neglected over the last decades may explain that most of the main studies in the development of the modern Linguistics omit in their pages the reference to this figure, of unquestionable scientific merits and of obligated reference when someone plans to work in the field of diachronic Linguistics.

Biography

Yakov Malkiel was born in Kiev (Ukraine) in 1914, to a Jewish family. Because of the civil war, he emigrated to Germany, where he studied at the University Friedrich-Wilhems in Berlin. In 1938, he obtained his Doctorate in Linguistic Romance degree with *suma cum laude*. Because he was jewish, he was forced to emigrate again in 1940, this time to the United States. Here, he worked for a short while at the University of Wyoming, and later he moved to the University of Berkeley (California), where he successively held the positions of *Lecturer*, *Assistant Professor* of Spanish and Portuguese, *Professor* of Romance Philology and *Professor Emeritus* of Linguistics and Romance Philology. In 1946, he founded the *Romance Philology* magazine, of which he was Editor in Chief for many years. In 1952, he participated in the creation of the Department of Linguistics of the University of Berkeley, where he carried out his educational activities until his death on April 24, 1998.

The merits of his career were recognized with numerous distinctions, such as three Guggenheim Awards and an *honoris causa* doctorate from the universities of Chicago (1966), Illinois (1969), Paris (1983), Berlin (1983), Georgetown (1987), Oxford (1989), and Salamanca (1994).

References

Malkiel, Yakov. 1958. Los interfijos hispánicos: Problema de lingüística histórica y estructural. Miscelánea homenaje a André Martinet. La Laguna: Universidad de La Laguna.

———. 1966. Genetic analysis of word formation. Current trends in linguistics. Vol. III. La Haya: Mouton.

———. 1968. Essays on linguistic themes. Berkeley: University of California Press.

———. 1982. Perspectives on historical Linguistics. Amsterdam/Philadelphia: John Benjamins.

———. 1983. From particular to general Linguistics. Essays 1965–1978. Amsterdam/Philadelphia: John Benjamins.

———. 1988. Yakov Malkiel: a tentative autobibliography. Berkeley: University of California Press.

———. 1989. Theory and practice of romance etymology: studies in language, culture and history. Variorum: Longdon.

———. 1990. Diachronic problems in phonosymbolism: edita and inedita, 1979–1988. Amsterdam/Philadelphia: John Benjamins.

———. 1992. Diachronic studies in lexicology, affixation, phonology: edita and inedita, 1979–1988. Amsterdam/Philadelphia: John Benjamins.

———. 1993. Etymology. Cambridge: Cambride University Press.

———. 1983. Introduction. In Malkiel. From particular to general Linguistics.

Rico, Francisco. 1969. Yakov Malkiel. Anuario de Estudios Medievales 6.

Hall, Robert A. 1987. Linguistics and pseudo-linguistics. Amsterdam/Philadelphia: John Benjamins.

Hymes, Dell, and John Fought. 1981. American structuralism. The Hague: Mouton.

JOSÉ CARLOS MARTÍN CAMACHO

See also **Etymology; Historical Linguistics; Indo-European 4: Romance; Language Change; Structuralism**

Manner of Articulation

Consonants are formed by creating a constriction in the vocal tract; the *manner of articulation* describes the type of constriction, whereas the *place of articulation* indicates where the constriction takes place. Sounds can also be made with different *airstream mechanisms.*

A primary aspect of the manner of articulation is the degree of the constriction, that is, whether the air has a free passage through the mouth or whether there is an obstruction. By varying the degree of obstruction, different types of sounds can be produced. Different sounds can also be made by allowing air to pass through the nasal passage; such sounds are known as *nasal.* Also, some sounds are *lateral,* with the air passing out through the sides of the vocal tract, but not through the center.

Vowels are produced with a very open vocal tract. For consonants, we distinguish three degrees of stricture: stops, fricatives, and approximants. *Obstruents* include the oral stops, fricatives, and affricates. Nasals and approximants are nonobstruent sounds, called *sonorants.*

Stops have a complete closure in the oral cavity. *Oral stops* involve closure of both the oral and nasal cavities, with no air passing through either one. *Nasal stops* have air passing out through the nasal cavity, but not through the oral cavity. In English, [p, t, k, b, d, and g] are oral stops, and [m, n, and ŋ (as in *long*)] are nasal stops. Nasal sounds are discussed further below.

Fricatives are made with an incomplete constriction in the oral cavity, allowing air to pass through the mouth, but with the constriction close enough to cause turbulence in the airstream, producing a hissing type of noise called *frication.* The English fricatives are [f v], [θ] as in *thick,* [ð] as in *the,* [s z], [ʃ] as in *shoe,* and [ʒ] as in *garage.* The fricatives [s z ʃ ʒ] are called *sibilants.* Fricatives are also known as *spirants.*

Affricates consist of a stop immediately followed by a fricative made at the same place of articulation. English has the affricates [tʃ] as in *itch* and [ʤ] as in *badge. Approximants* have an articulatory constriction that is narrower than the vowel [i], but not close enough to cause frication. Any approximant can be turned into a fricative by narrowing the constriction until frication begins. English has the approximants [l], [ɹ] as in *Mary,* [w] as in *wine,* and [j] as in *yes.* Some dialects of English pronounce, e.g. *which* with a voiceless approximant as well, symbolized as [ʍ]. Fricatives and approximants differ only in that approximants are slightly more open than fricatives. For this reason, where no approximant symbol is given on the International Phonetic Alphabet chart, an approximant can be shown by using the "lowered" diacritic [̞] with a fricative symbol; for example, a voiced bilabial fricative would be symbolized as [β̞], indicating that it is like a fricative, but with a more open vocal tract.

Two types of *trill* are commonly found. In one, the tip of the tongue is near the front teeth; in these, the tongue tip vibrates very quickly against the roof of the mouth several times. The other trill is uvular [ʀ]; with it, the uvula strikes against the dorsum of

the tongue. Snoring often involves an uvular trill made while breathing in.

Trills are not made by consciously controlling the motion of the tongue; rather, the tongue is placed in the appropriate position and tension, air is blown through the gap, and aerodynamic forces cause the tongue to vibrate rapidly against the roof of the mouth.

Taps are often described as a trill of one vibration. The tongue is flicked quickly against the roof of the mouth. A stop, by contrast, is a more deliberate action. In North American dialects of English, the tap [ɾ] is a typical variant of /t/, occurring between vowels, as in *city*.

Laterals are made with a central constriction, but with the sides of the constriction open, allowing air to escape. English has only the one lateral [l].

Nasals are made with a velic opening allowing air to pass out through the nasal passage. If the velum is raised against the pharyngeal wall (velic closure), the nasal cavity is shut off from the rest of the vocal tract, and no air can go through the nasal cavity. If the velum is lowered (velic opening), air can pass through the nasal cavity.

The term *nasal,* used alone, means *nasal stop.* Otherwise, the type of sound must be specified, as in *nasal fricative, nasal lateral,* etc. Nasal fricatives and approximants occur, but they are usually the result of being next to a nasal stop or vowel. Nasal stops are ordinarily voiced.

Air flowing out of the body is called *egressive.* Air entering the body is called *ingressive.* Most speech is egressive, but some sounds in some languages are made with an ingressive airstream.

In addition to the direction of airflow, there are three different airstream mechanisms that are used to set the air into motion (Chart 1). Most sounds are made by pushing air out of the body by contracting the lungs. These sounds are called p*ulmonic.* Pulmonic egressive stops are called *plosives.* Pulmonic ingressive sounds are not normally used in language (see below):

Sounds can be made by making an oral closure (such as a [p]), lowering the larynx, forming a glottal stop, and then raising the larynx. These actions cause

the air pressure to rise in the pharynx and oral cavity; on release of the oral closure, air flows out with a rather hollow, popping sound. This airstream mechanism is called *glottalic*; glottalic egressive stops are called *ejectives* [p' t' k']. Many native North American languages have ejectives.

Although it is possible to make glottalic ingressive stops, they are not at all common in languages. More often, we find *implosives* [ɓ ɗ ɠ]. These are made by raising the larynx, simultaneously making a stop, and then, as the oral stop is released, lowering the larynx so that the vocal folds vibrate as they move downward.

Clicks are made by making a velar closure as in [k] and simultaneously making an oral closure further forward (e.g. [p] or [t]). Then, the tongue body is pulled back, lowering the air pressure in the oral cavity. The release of the oral closure is a click; the airstream is *velaric ingressive.* The click symbols are bilabial [ʘ], dental [ǀ], alveolar [ǃ], palatoalveolar [ǂ], and alveolar lateral [ǁ]. As ordinary consonants, clicks are found only in languages of southern Africa; however, English speakers often use the dental click [ǀ] to express mild displeasure (often written *tsk! tsk!*), and the lateral click [ǁ] is used to urge on horses. Glottalic egressive sounds are not ordinarily used in language.

Further Reading

Catford, J.C. 1977. Fundamental problems in phonetics. Bloomington, IN: Indiana University Press.

Catford, J.C. 1988. A practical introduction to phonetics. Oxford: Oxford University Press.

Gimson, A.C. 1980. An introduction to the pronunciation of English. London: Edward Arnold.

Hardcastle, William J., and John Laver, (eds.) 1997. The handbook of phonetic sciences. Oxford: Blackwell.

International Phonetic Association. 1999. Handbook of the International Phonetic Association. Cambridge: Cambridge University Press.

Ladefoged, Peter. 1993. A course in phonetics. Ft. Worth, TX: Harcourt Brace Jovanovich,

Ladefoged, Peter. 2000. Vowels and consonants: an introduction to the sounds of languages. Oxford: Blackwell.

Ladefoged, Peter, and Ian Maddieson.1996. Sounds of the World's Languages. Oxford: Blackwell,

Laver, John. 1994. Principles of phonetics. Cambridge: Cambridge University Press.

Minifie, Fred, Thomas J. Hixon, and Frederick Hixon. 1973. Normal aspects of speech, hearing, and language. Englewood Cliffs, NJ: Prentice-Hall.

Pullum, G.K., and W.A. Ladusaw. 1996. Phonetic symbol guide. Chicago: University of Chicago Press,

Rogers, Henry. 2000. The sounds of language: an introduction to phonetics. Harlow, England: Pearson.

HENRY ROGERS

CHART 1

Stops	Pulmonic	Glottalic	Velaric
Egressive	Plosives p t k b d ɣ	Ejectives p' t' k'	——
Ingressive	——	Implosives ɓ ɗ ɠ	Clicks ʘ ǀ ǃ ǂ ǁ

Māori and the Polynesian Languages

Maori is the indigenous language of New Zealand, and since 1987, it has been an official language of the country. It was brought to New Zealand 800 to 1,000 years ago by voyagers from Central Polynesia, specifically from the Southern Cook Islands, and the area of modern French Polynesia. It is thus very closely related to the languages of these areas and less closely related to the other languages of the whole Polynesian family of languages. This language family, which forms the most easterly group within the Austronesian language family, contains some 30 languages, which are spoken by approximately 750,000 people in communities varying in size from more than 300,000 (Samoan) to just 100 or 200. The family encompasses all the indigenous languages spoken within the Polynesian Triangle, as well as many spoken within Micronesia and Melanesia. The Polynesian Triangle is a vast geographic area formed by lines joining New Zealand, Rapanui (Easter Island), and Hawaii, and the languages to be found here include Maori, Rarotongan, Tahitian, Rapanui, Hawaiian, Tongan, and Samoan. All the languages of this family are descended from Proto-Polynesian (PPN), a language spoken in the region of Tonga and Samoa some 2,500 years ago. As the Polynesian peoples dispersed and settled in the other regions of the Pacific, the language they carried with them diversified and, over time, became the modern

languages of this family. The languages named are related to one other as shown in the family tree, which shows the subgroups (subfamilies) within the family.

Through the comparative method, a great deal is known about the structure of PPN. The following table sets out examples of the data which allow the reconstruction of the sound system and vocabulary of PPN.

Tonga	Samoan	Hawaiian	Māori	PPN	English
pō	pō	pō	pō	*pō	'night'
valu	valu	walu	waru	*valu	'eight'
kai	ʻai	ʻai	kai	*kai	'eat'
tolu	tolu	kolu	toru	*tolu	'three'
fale	fale	hale	whare	*fale	'house'
fā	fala	hala	whara	*fara	'pandanus'
ʻā	ala	ala	ara	*ʔara	'awaken'
hala	sala	hala	hara	*sala	'wrong'
hala	ala	ala	ara	*hala	'way'
laʻā	lā	lā	rā	*laʔā	'sun'
ua	lua	lua	rua	*rua	'two'
fafo	fafo	waho	waho	*fafo	'outside'
fefine	fafine	wahine	wahine	*fafine	'woman'
moʻunga	mauga	mauna	maunga	*maʔuŋa	'mountain'

The asterisk indicates that the forms so marked are reconstructed by the methods of historical linguistics and not actually attested. The 'g' of Samoan and the 'ŋ' in PPN both designate the 'ng' sound of English 'sing'. The ʻ symbol in Tongan, Samoan, and

PPN

Tongan   Nuclear Polynesian

Samoan   Eastern Polynesian

Central Eastern Polynesian   Rapanui

Māori, Rarotongan, Tahitian, Hawaiian
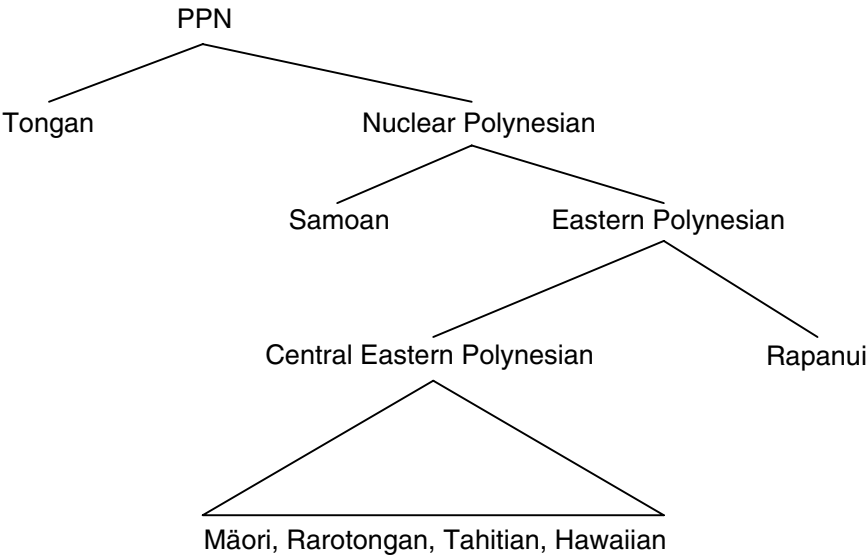
Hawaiian, as well as the ʔ in PPN, designate the glottal stop, often called a 'catch' or 'break,' articulated in the larynx. The macron over some vowels (*ā*) indicates that these are long.

An area of the grammar of PPN and of subsequent developments in its descendants that is of particular interest to scholars studying language typology is the matter of ergativity. Ergative languages, such as Tongan, mark the subjects of clauses differently depending on whether the verb is transitive or not, whereas nonergative languages, such as Maori, treat subjects in the same way, irrespective of the nature of the verb.

Tongan:

Na'e   'alu   e   tangata
Past   go   the   man
'The man went (intransitive verb 'go')'

Na'e   kai   'e   he tangata e   ika
Past   eat   subject marker   the man   the   fish
'The man ate the fish (transitive verb 'eat')'

Māori:

I   haere   te   tangata
Past   go   the   man
'The man went (intransitive verb 'go')'

I   kai   te   tangata   i   te   ika
Past   eat   the   man   object marker   the   fish
'The man ate the fish (transitive verb 'eat')'

Given that both patterns occur in the languages of the family, the question arises as to which of these patterns predominated in PPN and by what mechanisms the descendant languages developed the patterns presently attested (see especially Chung 1978).

The study of the PPN vocabulary and of the structure of the family produces hypotheses about the settlement of the Pacific that accord well with those derived from archeology. Polynesians entered the Triangle from eastern Fiji and settled the Tonga-Samoa area before a group moved further eastward to modern French Polynesia. From there, peripheral areas such as Rapanui, Hawaii, and New Zealand were discovered and settled.

Maori itself is, in terms of its sound system, grammar, and vocabulary, typical of the family as a whole. Maori has ten consonants, p, t, k, m, n, ŋ (spelt 'ng'), r, w, h, and f (spelt 'wh'), and five vowels, a, e, i, o, and u, which may be short or long. Furthermore, syllables are restricted in shape to at most one consonant followed by up to three vowels. Thus, in Maori, consonants occur only singly and never at the end of a word. Maori has very little in the way of inflection, that is, the use of devices such as suffixes for grammatical purposes. Rather, grammatical

categories such as tense and the role that words play within sentences are indicated by means of separate particles:

I haere 'went', *kua haere* 'has gone', *ka haere* 'will go, goes', *me haere* 'must go'.

Similarly, whether nouns are subjects, objects, possessors, and so on is marked by particles: in the Maori example above, 'The man ate the fish', the fact that 'the fish' is the object of 'eat' is shown by the particle *i*.

One area of Maori grammar in which a suffix plays a role is the formation of passive verbs:

horoi 'wash', *horoia* 'be washed'; *inu* 'drink', *inumia* 'be drunk (of liquids)'; *kai* 'eat', *kainga* 'be eaten'.

There are 17 such suffixes, although many of these are very rare. However, the question of which suffix goes with which verb and how this phenomenon is best described has been a subject of some discussion within theoretical linguistics (see especially Bauer 1997:477–9).

In Maori, as in most of the languages of the Polynesian Triangle, the predicate phrase (usually a verb) occurs first in sentences in which no special emphasis is placed on any one part. It is followed by the subject and any other phrases, such as objects and adverbial phrases. The examples given above illustrate this feature for both Maori and Tongan. These languages, however, also have devices for placing particular emphasis on some part, usually by placing the phrase in the initial position:

Nā   te   tangata   te   ika   i   kai
Of   the   man   the   fish   past   eat
'It was the man who ate the fish (not someone else)'

With the exception of some eight nouns, all designating people, which lengthen the third last vowel (*wahine* 'woman', *wāhine* 'women'), nouns in Maori show no change to form plural. Rather, number is shown in the accompanying articles:

te   whare   ngā   whare
the   house   the plural   house
'the house'   'the houses'

tērā   puke   ērā   puke
that   hill   those   hill
'that hill'   'those hills'

The Maori pronoun system shows two features typical of Polynesian languages:

- Separate forms not only for singular (one person) and plural (several people) but also for dual (two people): *koe* 'you (one)', *kōrua* 'you (two)', and *koutou* 'you (several)'.
- Separate forms for 'we' depending on whether the person spoken to is included (inclusive forms)

or not (exclusive forms): *tāua* 'you (one) and me', *māua* 'him or her and me', *tātou* 'you (two or more) and me', and *mātou* 'them and me'.

The possessive constructions in most Polynesian languages, including Maori, show two parallel sets of markers. Which is appropriate depends on the relationship between the possessor and the thing possessed, in particular, whether the possessor exercises independent control over the thing possessed:

tāku	*pukapuka*	but	*tōku*	*ingoa*
my	book		my	name
'my book'			'my name'	

Nā Hēmi tēnei kai	but	*Nō Pita tēnei whare*
Of James this food		Of Peter this house
'This food is James's'		'This house is Peter's'

Spread as it is over a country of the size of New Zealand, Maori shows some dialect diversity. This is, however, mostly a matter of vocabulary and is by no means so great that mutual intelligibility is impaired.

Once the language of the entire Maori population, it has been subjected to severe pressure from English, especially during the second half of the twentieth century. Of the approximately 520,000 people (15% of the New Zealand population) who identified themselves as ethnically Maori in the 1996 census, only some 26% claimed that they could conduct a conversation on everyday matters in Maori. Other studies have shown that there has probably been some overreporting by many respondents and that there may well be only as few as about 10,000 really fluent speakers of Maori. Further, the native speakers of Maori are concentrated in the older generations, and the natural transmission of the language from parents to children within families is essentially broken. In parallel to the decrease in the population speaking Maori, there has also been a reduction in the domains where it is regularly used. In most parts of the country, Maori is regularly used only during ceremonial occasions, such as the welcoming of guests and at funerals and church services.

The awareness that the continued survival of Maori had become precarious and the strongly held feeling that the language is a central feature of Maori ethnic and cultural identity have led, since the mid-1970s, to a wide range of initiatives in language maintenance and development. In response to the recognition that the natural transmission of the language could no longer be relied on to ensure its acquisition by succeeding generations, preschools were founded in which Maori was the sole language spoken and in which instruction took place within a Maori cultural environment. Since the early 1980s, this initiative has extended into primary and secondary schooling, catering to approximately 20,000 pupils from preschool to high school levels. Maori Studies is a popular subject at all New Zealand Universities and can be pursued up to the doctoral level.

In 1987, the passage of the Maori Language Act made Maori the only de jure official language of New Zealand and founded the Maori Language Commission. This body, with its office and staff in Wellington, develops policy and promotes activities designed to give substance to the official status of the language. Part of the Commission's work has centred on the development of vocabulary. Maori is now used in official documents, in broadcasting, and in education as never before, and deliberate expansion of its vocabulary to equip it for these domains has been essential. In developing new vocabulary, the Commission and other authorities have been guided by the principle that terms should be created with the resources already found in the language. Thus, borrowing from English is ruled out, and the creation of new compounds or the specialisation of existing words are preferred strategies.

Maori is the vehicle of a considerable traditional literature. Oral traditions of a variety of genres, such as settlement histories, aetiological myths, fairy stories, songs, and poetry, were committed to writing by Maori elders themselves and are accessible partly in publications and partly in manuscripts in libraries and museums. Oratory and song are two oral genres that are still very much alive today and play central roles in Maori culture and community life. At the same time, creative writing, especially for school use, is using Maori within new literary genres, such as the short story and even the novel.

The maintenance and development of Maori is a very important strand in New Zealand's present movement toward a more equitable, multicultural, and inclusive social structure, as past wrongs and assimilationist policies are redressed.

References

Bauer, Winifred. 1997. The Reed reference grammar of Māori. Auckland: Reed.

Biggs, Bruce. 1963–1976. The languages of Polynesia. Current trends in linguistics, Vol. 8, ed. by T. Sebeok, 466–505. The Hague: Mouton.

Biggs, Bruce. 1969. Let's learn Maori. Wellington: Reed.

Chung, Sandra. 1978. Case marking and grammatical relations in Polynesian. Austin: University of Texas Press.

Krupa, Victor. 1982. The Polynesian languages. London: Routledge & Kegan Paul.

Lynch, John. 1998. Pacific languages an introduction. Honolulu: University of Hawai'i Press.

Williams, H. W. 1971. A dictionary of the Maori language, 7th edition. Wellington: Government Printer.

See also **Austronesian; Comparative Method; Endangered Languages; Genetic Relationship; Language Planning; Typology**

Marshallese and Micronesian Languages

Malayo-Polynesian is one branch of the Austronesian language family, and has two sub-branches: Western Malayo-Polynesian (which includes a large number of languages, including some of the major languages of the family) and Central-Eastern. The Central-Eastern sub-branch in turn consists of two major branches: Central Malayo-Polynesian and Eastern Malayo-Polynesian. The Eastern Malayo-Polynesian in turn is subdivided into two sub-branches: South Halmahera-West New Guinea and Oceanic. The Oceanic sub-branch contains over 450 languages, or about half of the total for the Austronesian family as a whole. Most Oceanic languages have lost word-final consonants and have simplified consonant clusters. The most important sub-branch, Remote Oceanic, has a number of branches and sub-branches. The more important groups of languages belonging to Remote Oceanic are Micronesian and Central New Hebrides. The Remote Oceanic sub-branch covers most of Micronesia and all of Polynesia (including Fijian, Rotuman, Samoan, Maori, Tongan, Tahitian, Hawaiian, and Rapanui or Pascuense). However, a number of scholars question the validity of the Remote Oceanic subgroup.

The languages of most of the Micronesian islands (except the languages of Palau and the Marianas) are more closely related to those of eastern Melanesia (the Solomon Islands and Vanuatu). Linguistic analysis points toward eastern Melanesian origins for the earliest eastern Micronesian peoples. The Micronesian subgroup of the Remote Oceanic sub-branch includes all the languages of Micronesia except for Palauan, Yapese, and Chamorro (which all belong to the Western Malayo-Polynesian branch and have affinities with the western Austronesian languages of Indonesia and the Philippines) and Nukuoro and Kapingama-rangi (which are Polynesian). There are approximately 73,000 total speakers of Chamorro, of whom 60,000 are on guam and 13,500 on the Northern Mariana Islands, Micronesia. The 15,000 speakers of Palauan or Palau primarily reside in Belau, guam, and Western Carolines, Micronesia. And there are roughly 5,000 speakers of Yapese on Yap, Caroline Islands,

Micronesia. Many scholars question the position of Yapese in this group since no one has yet demonstrated that it must belong to this group. Several scholars, including Blust, have recently proposed that this language belongs to the Oceanic subgroup. The approximately 1,300 speakers of Kapingama-rangi include 700 on Kapingama-rangi and several hundred in Porakiet village on Ponape. It is somewhat intelligible with Nukuoro, with 55% lexical similarity. There are roughly 550 speakers of Nukuoro, with 125 on Ponape. The chief languages of the Remote Oceanic sub-branch include: Ikiribati (Gilbertese) with over 60,000 speakers, mainly in Kiribati; Trukese with about 45,000 speakers, including second-language speakers; and Marshallese (over 30,000 speakers).

Geographically, the Micronesian Islands lie within four major archipelagos: the Marianas, the Carolines, the Marshalls, and the Gilberts. There are 19 languages in the Micronesian subgroup including Carolinian (Saipan Carolinian), English, Kapingamarangi, Kusaie (Kosrae, Kosraean, Kusaiean), Mokil (Mokilese), Mortlock (Mortlockese), Namonuito, Ngatik, Nukuoro (Nukoro, Nuguor), Paafang, Pingelap (Pingilapese), Ponape (Pohnpeian, Ponapean), Puluwat, Satawal (Satawalese), Sonsorol, Truk (Ruk, Trukese), Ulithi, Woleain (Wolean), and Yapese. Kiribati is the language of the Kiribati (Gilbert) island chain, of neighboring Banaba (Ocean) island, Niu, and it is also spoken in the Line Islands. Kusaie is the language of 5,000 inhabitants of Kosrae State (Federated States of Micronesia), and many inhabitants also speak Ponapean. Nauruan is the traditional language of Nauru, with approximately 4,000 speakers. English has been the strongest influence on Nauruan (Australia administered the island from 1920 to 1968), and there are signs that Nauruan is being abandoned in favor of English. Ponape, with about 22,000 speakers, is the major language of Pohnpei State (Federated State of Micronesia). Ngatikese, Pingelapese, and Mokilese are counted as dialects of Pohnpeian or as closely related languages. Ponape has a 'high language' with a

partly separate vocabulary, used in speaking about people of high rank. Truk (22,000 speakers) is the major member of a dialect chain with a total of 40,000 speakers. These languages are spoken in the Caroline Islands (Federated States of Micronesia), and have numerous English and Japanese loanwords from their colonial history.

Marshallese has the largest number of speakers (about 30,000) of any Micronesian language. Marshallese is primarily spoken in the Republic of the Marshall Islands, but also on Nauru. English and Japanese have had the strongest effect on Marshallese. Japanese administration followed after World War I, and the United States took over from 1945 to 1986.

The Marshall Islands comprise 29 coral atolls and five single coral islands scattered over about 180,000 square miles of the Pacific Ocean. Kwajalein Atoll is the world's largest atoll formation, comprising 93 islets. Considering this vast area and scattered land area, the Marshallese language is amazingly homogenous, but there are two main dialects that sometimes differ in vocabulary, the choice or the pronunciation of a word, and in the vowels of words whose first two consonants are identical. These two dialects coincide with the two major chains of islands: the eastern, sunrise, or Ratak chain and the western, sunset, or Ralik chain. However, at focal centers like Majuro, or Ebeye on Kwajalein, one finds speakers of both dialects. Zewen (1977) found evidence of the two dialects merging into a kind of Marshallene 'koine' in the District Centre, as well as on Ebeye.

The Ralik dialect has obtained a higher status because the Bible was first translated into Ralik, and the early missions were located in that chain. The eastern (Ratak) dialect is more conservative. The differences between the two dialects appear in the vocabulary and the phonetic realization, rather than in the grammar. For instance, words that begin with a vowel have a different phonetic onset in the two dialects. The Ratak dialect has quite a number of words that consist of a reduplicated basic morpheme. There are also minor dialects within the major dialects: Mejit, for example, is divergent in the east, as in Ujelang in the west.

Terms associated with precolonial customs and activities such as native clothing, handicraft, and navigation are today referred to as 'Kajin Etto' (old language), although understanding of these terms varies in the present generation (Zewen 1977:4–6). In precolonial times, ritual languages were mandatory for certain activities: for instance, special words had to be used on the northern voyages, especially to Rongerik. Word taboos were associated with tattooing or particular places: for instance, on the small islands of Kwajalein Atoll, 'kako' (rooster) was taboo. Most terms of these ritual languages are metaphors or metonymies.

Students and youngsters have also developed a 'student language' or slang in Majuro and on Ebeye (Zewen 1977). It is characterized by nonstandard vocabulary, arbitrary coinages, and figures of speech that often quickly fall out of fashion or use.

Linguists have differed over the vowels of Marshallese. Certainly, Marshallese has a very high number of vowel phonemes, particularly in comparison with Ponapean (with which it has roughly 33% lexical similarity). This situation seems to be the result of vowel fusions after certain consonants have disappeared according to Marshallese sound laws. Byron Bender (1968) argues that there were three vowel phonemes (*a e I*), other vowel sounds being conditioned by surrounding consonants. The usual spelling recognizes nine. Other scholars distinguish 12 vowel phonemes, and add a distinctive feature of length, giving a total of 24. There are three groups of consonant sounds, according to their effect on neighboring vowels: the plain or light consonants, the heavy consonants, and the rounded consonants. Marshallese has lost the complex series of numeral classifiers that are found in the other Micronesian languages. 'Cilcino' (six) originates as 'cilu' (three) plus 'cilu' (three); '7' as '3 + 3 + 1'.

With respect to Marshallese phonology, there are eight basic patterns of syllables. Words, when monosyllabic, consist of one of these patterns; when polysyllabic, they consist of a sequence of these patterns. Most Marshallese words are dissyllabic.

With respect to morphology, Marshallese is not an inflected language in which word classes can be defined by the type of inflections they take. Other criteria have to be considered such as the meaning of the word, and the valence or systematic combinability with other words. For instance, nouns may be determined by demonstrative pronouns. They may be main or secondary elements of phases or may be preceded by a preposition. In clauses they may be subject, predicate, or object. With the exception of personal and nonpersonal proper nouns, they may take 'possessive suffixes' or may be constructed with possessive classifiers.

References

Abo, T., B.W. Bender, A. Capelle, and T. DeBrum. 1976. Marshallese–English dictionary. Honolulu: University Press of Hawaii.

Bender, B.W. (ed.) 1984. Studies in Micronesian linguistics.

Bender, B.W. 1968. Marshallese phonology. Oceanic Linguistics 7. 16–35.

Bender, B.W. 1984. Object marking in Marshallese. In Bender, pp. 443–65.

Blust, R. 1990. Summary report: linguistic change and reconstruction methodology in the Austronesian language family. Linguistic change and reconstruction methodology, ed. by P. Baldi, 133–53. New York: Mouton de Gruyter.

Clark, R. 1987. Austronesian languages. The world's major languages, ed. by B. Comrie, 899–912. New York: Oxford University Press.

Combs, M.R. 1984. Glides: the problem of definition and analysis, including a discussion of the treatment of semivowel glides in Micronesia and an acoustic phonetic analysis of glides in selected Trukic dialects. Ph.D. Thesis. Honolulu: University of Hawaii.

Kroeber, A.L. 1911. Phonetics of the Micronesian language of the Marshall Islands. American Anthropologist 13. 380–93.

Lingenfelter, S.G. 1986. Social anthropology and translation. Notes on Translation 116. 30–46.

Matsuoka, S. 1929. Study of the language of the Marshall Islands. Tokyo: Kyodo Kenkyusha.

Rehg, K. 1984. The origins of 'compensatory lengthening' rules in Micronesian languages. In Bender, pp. 53–59.

Zewen, F.-X. 1977. The Marshallese language: a study of its phonology, morphology, and syntax. Unpublished Ph.D. Dissertation. Hamburg: Universitat Hamburg.

CAMILLA COCKERTON

Martinet, André

The main representative of functionalist structuralism, André Martinet is one of the most enlightening figures of twentieth-century linguistics. He spent his childhood in the peripheral Provencal region of Savoy, which endowed him with an early sensitivity for the problems of bilingualism and language social aspects, while during his brilliant classical education he acquired practical knowledge of a number of languages. His degree was in Germanic linguistics but he also acquired broad knowledge of the general European linguistic and philological tradition. During his education in Paris, Berlin, and Copenhagen, as well as later while he was teaching at Columbia University, he established direct contacts with all relevant structuralist trends in both Europe and America (developed by the Prague School, the Copenhagen School, and American structuralism), showing appreciation of all elements of other schools that could fit his own view of language. Influenced primarily by the structuralism of Ferdinand de Saussure and that of the Linguistic Circle of Prague (particularly Nikolay Trubezkoy), Martinet further developed the structuralist theoretical tenets enriching them by achievements of American linguistics and Danish glossematic theory. However, adopting a clearly functional view, he specifically emphasizes the functions of language in general, or the functions of linguistic elements as the main object of linguistic enquiry in opposition to structuralism and formalism of the majority of American structural linguists in which the system of a language is explicitly studied in abstraction from its functions. For him, language is essentially a human institution and an instrument of communication, while the method of linguistic analysis is based on the principle of communicative relevancy of linguistic units and their role in the transmission of information. Accordingly, he sees language structure only as a logical complement and manifestation of language functioning.

Within this functional perspective, Martinet clearly insists on the reliance on concrete linguistic facts and the need to study specific languages exclusively in their own functional framework, as each language corresponds to a specific organization of reality and human experience: 'What we wish to examine is the whole of man's linguistic activity, and as this activity only takes place within the framework of specific languages, our primary task is to study these languages as so many different objects.' This rational and prudent attitude, which is in sharp contrast to perhaps more attractive, bold theoretical assumptions of some other formalistic and structural linguists, has often been labeled as Martinet's 'realism'. Martinet maintains that the task of linguistic analysis is not to seek manifestations of universal categories in languages, but to discover the unique categories articulated by each language. Accordingly, in his view, theory should be always guided by analysis, not the other way around, no matter how unfamiliar the resulting theory might appear, as 'the scientific treatment of an object requires, first and foremost, that the object's integrity should not be sacrificed to methodological exigencies.'

A cornerstone of the functionalist approach and Martinet's main contribution to general linguistics is his important notion of 'double articulation' referring to two fundamental levels of language structure, that of meaningful units (primary articulation) and that of expression units (sounds) which are meaningless in themselves but serve to distinguish the meaningful elements (secondary articulation). The concept of 'double articulation' (or 'duality of patterning' as it is termed in American linguistics to avoid confusion with

the articulation of sounds) allows a small number of meaningless phonological elements to be combined to distinguish an infinite set of meaningful utterances. The minimal linguistic forms that are meaningful are monemes, corresponding roughly to what American linguists call morphemes, which Martinet further analyzes as units of grammatical meaning (morphemes) and units of lexical meaning (lexemes). Thus, the word *books* is composed of two monemes: a lexeme *book* and a morpheme *-s* denoting plurality. Moneme has both an expression and content and consists in the association between a signifier (vocal expression) and a signified (semantic content). In line with the structuralist tradition of the phonological Prague School the basic unit of the sound system is called a phoneme: it is a minimal, contrastive sound unit that distinguishes one meaningful utterance from another (e.g. *pin/bin*). Phonemes may be further analyzed in terms of a set of underlying distinctive features specifying the ways the sound is physically produced. However, Martinet rejects Roman Jakobson's phonemic theory that there is a universal inventory of these distinctive features and that they are based on binary oppositions.

Martinet made major contributions to the field of general and descriptive phonology further developing the phonemic theory of the Prague School through elaboration of such notions as phonemic frequency, functional load, neutralization, and segmentation of phonemes. His second doctoral dissertation *La phonologie du mot en danois* (1937; Phonology of Danish) was one of the first complete phonological descriptions of a language. He regards phonology as 'functional phonetics', which has to interpret real phonetic facts from the point of view of linguistic functions of sound differences and their functional effects and not as different realizations of an 'abstract system'.

Using the function of moneme in the utterance as the starting point, Martinet also developed a coherent system of general morphological and syntactic analysis focusing particularly on the sentence structure and predicate function. While rejecting logical and psychological criteria, he analyzes sentences exclusively by applying the criterion of linguistic function (revealing linguistic 'choices' of speakers by means of commutation). He establishes a hierarchical relationship between monemes, the minimal units of the first articulation according to their function in the sentence by classifying them into free monemes and bound or functional monemes. The word can consist of one nonanalyzable item or moneme, e.g. *head* and of combined items or synthemes such as *headache*. In terms of communicative relevancy, morphology is considered to be the 'presentation of the variations of the form of signifiers and of the conditioning of these variations', and syntax is 'loosely understood as the survey of all data pertaining to the combinations of significant elements'. Syntax is not completely predictable but dependent on compatibility or conventions, which one just has to know and cannot calculate in advance. Martinet, however, has been criticized for avoiding to a certain extent the problem of semantic analysis by arguing that all linguistically relevant aspects of meaning must already be manifested at the level of expression, while its other aspects easily escape objective analysis and require nonlinguistic (external) explanation and therefore should not be the subject matter of linguistics.

An important part of Martinet's work is devoted to historical linguistics and diachronic problems of language change. In his seminal work *Économie des changements phonétiques* (1955; Economy of sound changes) he formulates an original, well-grounded theory of sound change as arising, in part, from internal pressures within a phonological system. He explains that linguistic evolution is 'governed by the permanent conflict between man's communicative needs' (using as many units as possible, as different from each other as possible) and 'his tendency to reduce to a minimum his mental and physical activity' (using as few units as possible). By economy, Martinet means the principle of 'least effort' and the tendency to reconcile these two contrasting elements to attain an improved functional load and help maintain an acceptable level of communication. However, optimum economy and complete symmetry can never be reached because of physiological factors such as inertia and asymmetry of the speech organs (e.g. distinction between front and back articulation). Martinet distinguishes syntagmatic economy (contextual assimilating pressure exerted by adjacent linguistic units) from paradigmatic economy (dissimilating pressure exerted by the units in the system which might have appeared in that particular place), and determines the effect of different variable synchronic factors (such as frequency, correlative relationships, the notion of empty place in a pattern, etc.) on language development. In this way, he tried to reconcile the apparently conflicting viewpoints of synchronic and diachronic linguistics by drawing upon his findings in one field to illuminate the other. However, he has been often criticized on account of deliberate neglect of external historical and social aspects of language change. Although it is true that Martinet primarily considered the inner causes of language change, he also underlined the role of external explanation in the rate of change and its specific historical context, and in this way has convincingly succeeded in elucidating many mechanisms of sound change making his work indispensable in any serious study on language evolution.

In addition to his important theoretical achievements, Martinet also wrote about specific problems of many different languages frequently discussing methodological and practical questions of language learning and acquisition, while at the same time working on the development of international auxiliary languages. With his central position between European and American structuralist schools and explicit appreciation of positive outcomes of traditional comparative philology, Martinet was one of the rare linguists who tended to bridge the gap between different fields of linguistic study. His clearly formulated functional theory based on the empirical-deductive approach exerted an important influence on further development of European linguistics, while his method, which always remains close to the actual reality of specific languages, has been used for phonological and syntactical descriptions of more than hundred languages.

Biography

André Martinet was born in Saint-Albans-des-Villards, Savoy on April 12, 1908. He obtained degree in Germanic linguistics from Université de Sorbonne and Ecolé Pratique des Hautes Etudés, Paris (1932), and a Ph. D. for two dissertations in Germanic linguistics on consonant gemmination in Germanic languages and on phonology of Danish in 1937. He was Chair of Phonology, Ecolé Pratique des Hautes Etudés, Paris in 1938–1946. During World War II, while imprisoned in German camp, he carried out an impressive phonetic and phonological survey on contemporary French pronunciation published in 1945. He moved to New York, where he was director of interlinguistic research of International Auxiliary Language Association (IALA) 1946–1948. He was Professor at Columbia University and was an active member of the Linguistic Circle of New York 1948–1955. He was Chair of Phonology, Ecolé Pratique des Hautes Etudés and Professor of general linguistics at Sorbonne, Paris in 1955–1978; member,

Société de linguistique de Paris in 1931; President, Société européenne de linguistique, 1966–1999; honorary member, Linguistic Society of America (LSA) in 1973, Union Mundial pro Interlingua (UMI) in 1998; editor of *Word* 1946–1960; founder and editor of *La Linguistique* since1965. Martinet died in Châtenay-Malabry on July 16 1999.

References

Fried, Vilém and René Dirven (eds.) 1987. Functionalism in linguistics. Amsterdam: Benjamins.

Fuchs, Catherine, and Pierre Le Goffic. 1992. Les linguistiques contemporaines: repères théoriques (Current linguistics: theoretical standpoints). Paris: Hachette.

Lepschy, Giulio C. 1970. A survey of structural linguistics. London: Faber and Faber.

Leroy, Maurice. 1963. Les grands courants de la linguistique moderne, Brussels: University Press of Brussels; as The main trends in modern linguistics, transl. by Glanville. Price. Oxford: Basil Blackwell, 1967.

Mahmoudian, Mortéza. (ed.) 1979. Linguistique fonctionnelle: débats et perspectives. (Functional linguistics: discussions and perspectives). Paris: PUF.

Martinet, André. 1955. Économie des changements phonétiques: Traité de phonologie diachronique (Economy of sound changes: Diachronic phonology), Bern: Francke Verlag.

———. 1960. Eléments de linguistique générale, Paris: Armand Colin; as Elements of general linguistics, transl. by Elizabeth Palmer. London: Faber and Faber, 1964; Chicago: University of Chicago Press, 1982.

———. 1962. A functional view of language. Oxford: Clarendon.

———. 1975. Studies in functional syntax (Études de syntaxe fonctionnelle). Munich: W.Finck.

———. 1989. Fonction et dynamique des langues (Function and dynamics of languages). Paris: Armand Colin.

———. 1993. Mémoires d'un linguiste. Vivre les langues (Memoirs of a linguist. Living the languages). Paris: Quai Voltaire.

Sampson, Geoffrey. 1980. Schools of linguistics, Stanford, CA: Stanford University Press.

Tobin, Yishai. 1988. Phonetics versus phonology: the Prague school and beyond. The Prague School and its legacy, ed. by Yishai Tobin. Amsterdam and Philadelphia, PA: Benjamins.

ANITA SUJOLDZIC

Mass Media and Language

Broadcasting (TV, radio), print media (newspapers, magazines, books, pamphlets, posters), film, theater, and electronic products such as CDs, records, tapes, and videos are the most important instances of mass media. Telecommunications and news agencies, the two other areas of mass communications, differ from them in a number of ways, but the boundaries are being blurred by modern, especially digital, information

technologies. Important characteristics that they all share are the unidirectionality of the information flow from a source to a target and the constraints they impose on feedback. Broadcast and print media have developed response channels, which differ in kind, immediacy, and effect as a result of the technology they are based on. They are often referred to as news media since they normally carry news, current affairs, and sports. But news need not be a defining criterion as music, film, or weather channels show. Although they will be ignored, one should add that the linguistic effects of entertainment media may surpass those of the informational and educational media and may further endanger aid endangered languages.

Before turning to linguistic aspects, one must emphasize the fact that they are (sociopolitical) institutions with, as a special characteristic that emanates from their purpose, developed mechanisms for the flow, categorization and packaging of information, and a history of professionalism. In terms of communicational function, they can thus be defined as domains whose goal it is to make content public, accessible to debate. But they are better referred to as hyperdomains since all domains, e.g. politics, law, business, lifestyle, are providers of content. Their sociopolitical role, especially their relationship with sources, and their impact on audiences remain controversial. What is clear is that they are not mere transmitters of information; they may indeed have a symbiotic relationship with sources (e.g. in sport). Minimally, they should be called co-orchestrators of what is 'in the air' in the public arena.

Turning to language, co-orchestration means that the media collect, select, manipulate, shape content, and create a packaged, marketable product that is recognizable to large audiences and helps maintain a media outlet's audience share. Language is the central, but not the only, component in these processes, the others being visuals, film, sound effects, presentation, and layout. Language is crucial in the packaging of *raw* input into output and in mediating it to an audience. The inference that media language must be homogeneous, however, is unwarranted in light of the diversity of source domains, professional routines, technologies (print, radio, or TV), audience stratification and habits, and the need for outlets to create recognizable and consumable products. It is better to speak of media languages or to borrow from media sociology the term public idiom, which is better able to refer to the existence of abstract norms and professional routines that bear upon the use of language. The linguistic concept of register has been used to describe such differences as they were seen to emanate from a particular configuration of situational features or context-of-situation. Early definitions of this term highlighted lexicogrammatical features or

clusters of co-occurring features. Recent ones define it in terms of semantic potentials, the typical meanings expressed in such contents, and the typical linguistic expressions used to do that end. Expressions range from text or discourse properties to features of pronunciation. Such a wide definition is to better able to capture the diverse means of expression that are used in restricted reported domains to shape content and to express underlying ideological stances.

Focusing more narrowly on linguistics, one should begin by saying that, in view of the diversity of linguistic models, it is functional ones, pragmatics, text and discourse linguistics, socio-and psycholinguistics that have an affinity to the concerns of media studies. Four points should be emphasized. The first is that linguistics can be an ancillary discipline or an independent player. In the latter capacity, it pursues its own agenda, using media output as primary data for the descriptive study of language. In the former, it helps highlight aspects of the communicational process that are of interest to media studies. Second, media output does not occur in isolated units but as instances of higher order units, such as program formats, or as products of a medium's institutional structures, such as news and current affairs. Output is thus subject to ever more abstract norms and expectations. Third, output has a presentative and representative side. The former highlights style or, to use Hallidayan terms, the 'interpersonal' function, the latter the way reality is construed, the 'ideational' function. In other words, presentation is about the way media output targets its audience, representation about the way it (re-)constructs what it is about. Both presentation and representation are supported by ideologies and attitudes that they reaffirm. Finally, one must distinguish output as a product, the result of complex institutional processes and output as a process, the way it is experienced by an audience. Linguistics can shed light on both, including the points mentioned above, viz. the embeddedness of the public idiom inside a network of source and target domains.

Studies on sports coverage have revealed that the technology, i.e. radio, TV, or print, and the mode of communication, i.e. speech or writing, have effects on the idiom, which are independent of such factors as program formats. Radio, for instance, turned out to be more action-oriented, focusing more on detailed spatial expressions than TV and the press, even though the time constraints on radio commenting are much heavier. Combining discourse and text linguistics with Content Analysis, studies on representation have left no doubt about the existence of ideological bias in media coverage. They have revealed mechanisms whereby 'hidden meanings' are transmitted that reinforce attitudes such as racism, in the absence of an audience's direct experience with ethnic groups. Studies closer to

mainstream linguistics have looked at how language is used to recast reality into a socially coherent consensus or, less often, conflictual position. 'Critical Linguistics' or Critical Discourse Analysis, outgrowths of Hallidayan functional grammar, have thus revealed how lexical fields, grammatical systems like transitivity or modality can be used to create ideological bias. Gender analyses can be seen in this context. Text and discourse analysis have looked at macro-and microproperties of media texts. Van Dijk argues forcefully that texts express, or can be reduced to, macrosemantic propositions and are condensed in conventionalized (text-) syntactic super-structures. He exemplified his approach in a study on the international press coverage of the assassination of Lebanon's President General Gemayel, which revealed the presence of a bias in favor of 'Western' formulations of the event and the dominance of local (as against global) journalistic routines. They showed up in the shared macrostructure of media texts. From a methodological perspective, he showed how (under-lying, social) ideologies and (concrete acts of) discourse are connected through (cognitive) models that discourses such as editorials aim to activate to bring over the writer's intent. Studies on the relationship of a text's structure with news values and narrative theory have shown that news values take precedence over the chronology of events. Reports, for instance, take the form of an inverted pyramid of information structure, where the most important, often the most recent, information comes first, with background being delayed. This anticlimactic structure can be traced to the American Civil War, where the emphasis was placed on 'recency' as against 'explanatory background'. It follows that reports do not need to be read to the end. Editorials, in contrast, which argue for or against some position, have a climactic structure and require full-text reading.

A traditional concern of presentation studies has been with the choice of language (variety). Choice was seen to respond to such factors as program policy, intelligibility, and to favor standard or prestige languages. This turned out to be true only in specific situations, as when public service media enjoy a monopolistic position, when broadcasting focuses on high cultural products or on nation-building. Britain's BBC between the 1920s and 1950s is a case in point. Competitive settings as in the United States of America or federal ones as in Germany promote a wider range of varieties. It has also been shown that standard varieties reflect the tone of officialdom and of elites, and support representational bias. A strict separation between presentation and representation is thus unrealistic and generalizations about either must be tempered in light of differences between intended norms and practices, and between intended communicative effects and its actual realization.

Generally speaking, research has moved away from narrow linguistic concerns (grammatical structures, lexical choices) to broad concepts that permit the understanding of ideological bias. It is here that the audience dimension has been reintegrated, whereas older research was largely confined to comprehension and recall studies. Like all texts, media texts are assumed to contain slots for the reader, which activate in recipients shared concepts, schemata, prototypes, and the like. Mass media, thus, do not target passive recipients; they activate recipients' worldview. Although this is an attractive position, the implication is that mass media address individuals, while media sociology rightly operates with layers of recipients that reinforce or cushion off media impact. Related research has looked at linguistic details to show how texts incorporate features of the language of the audience ('audience design') to enhance their representational impact and to maintain or increase audience size.

Studies of interviews, talkback shows, etc., have revealed patterns similar to the ones in monologs. Research tends to highlight the product—messages that recipients are most directly exposed to—although they interpret findings by relating them to institutional or general social parameters. Studies of the institu-tional context of the public idiom, i.e. the formulation of communicative and language norms, reveal the institutions' internal structures, such as the contribution of advisory bodies, language guardians, or of editorial policies. To turn to linguistics as an independent player, one should note the traditional interest of descriptive and applied linguistics in media language to study the 'real', as against the 'potential', in language use. The design of reference and teaching materials have been important outcomes. Corpus linguistics has added a new methodological dimension to comparative research.

Although linguistics has important contributions to make to mass media studies, one should not forget that factors to do with ownership, legal status, business strategies in light of the globalization of communi-cation, political control, public accountability, etc., dominate research.

References

Bell, Allan, and Peter Garrett (eds.) 1998. Media discourse. Oxford: Blackwell.

Burger, Harald. 1990. Sprache der Massenmedien. Berlin: Walter de Gruyter.

Dijk, Teun A. van. 1988. News as discourse. Hillsdale, NJ: Lawrence Erlbaum.

———. 1998. Opinions and ideologies in the press. Media discourse, ed. by Allan Bell, and Peter Garrett, 21–63 Oxford: Blackwell.

Halliday, Michael A.K., and Ruqaiya Hasan. 1976. Cohesion in English. London: Longman.

Fowler, Roger. 1991. Language in the news. Discourse and ideology in the press. London: Routledge.

Leitner, Gerhard. 1997. The sociolinguistics of communication media. Handbook of sociolinguistics, ed. by Florian Coulmas, 187–204. Oxford: Blackwell.

Leitner, Gerhard. 2000. Lexical frequencies in a 300 million word corpus of Australian newspapers: analysis and interpretation. International Journal of Corpus Linguistics 5(2). 1–32.

Leitner, Gerhard, and Markus Hesselmann. 1996. 'What do you do with a ball in soccer?' Medium, mode and pluricentricity in soccer reporting. World Englishes 15(1). 83–102.

GERHARD LEITNER

Mathematical Linguistics

Mathematical linguistics is a branch of linguistics studying the features and properties of language from the mathematical point of view. It is usually divided into several branches: (1) quantitative or statistical linguistics, dealing with unique language entities on the phonetical, lexical, syntactical, and semantical levels; (2) algebraic linguistics; and (3) machine translation. Algebraic linguistics use qualitative mathematical methods like: algebraic methods, theory of algorithms, graph theory, topology, and mathematical logic. Mathematical linguistics develops formal apparatus for the description of natural and artificial languages. The discipline was created in the 1950s to define language phenomena logically, precisely, and exactly. Its fundamental based on the ideas that language is a functional mechanism, which can be studied through the frequency of occurrences of speech units. More often than not, under mathematical linguistics, its statistical branch is understood. Mathematical linguistics unites the study of quality and quantity in language covering such fields as phonological, lexical, grammatical, semantic, and stylistic statistics. The frequency of occurrence of linguistic units and features, their rank, order, laws of distribution, correlation, stability, and variance coefficients are studied. Considerable attention is given to the definition of the units of population, random sampling, necessary, and sufficient sample size. The forms of frequency distributions of linguistic units are studied.

Phonological statistics studies the frequency of occurrence of letters, phonemes, and phonemic groups in languages as it is and in different languages world over. One can establish the phonostatistical distances between the sound pictures of different writers or different world languages. In 1937, a Japanese linguist Masao Onishi was one of the first to compare the frequency of occurrence of groups vowels and consonants of ten languages: English, German, French, Spanish, Italian, Russian, Chinese (Cantonese), Bengali, Sechuana, and Japanese. Yuri Tambovtsev compared the frequency of the occurrence of labial, front, palatal, sonorant, occlusive, fricative, and voiced consonants in 109 world languages of different language families, e.g. Indo-European, Finno-Ugric, Samoyedic, Tungus-Manchurian, Turkic, Paleo-Asiatic, etc. He built phonostatistical distances between the languages inside each language group (e.g. Germanic, Romanic, Slavonic, Ugric, etc.) and each language family.

Lexical statistics investigates the sizes of corpus, the zones of words with high, medium, and low frequencies, the laws of their distribution, especially Zipf law. Lexical statistics gave rise to glottochronology, created by an American linguist Mooris Swadesh in the 1950s. He tried to make statistical comparisons between the rates of change within the fixed list of words of two languages to find out the time these two languages separated. The glottochronological method is based on two pivotal assumptions: (1) in every language, there exists a list of 100 core words; (2) the rate of disappearing words is constant. This method was much criticized. The main methodological fault lies in the fact that it is hard to establish the main 100 word core of a language and the false relation of the decay of carbon-14 with the disuse of words in a language. Nevertheless, if one takes the whole word stock of two languages, then one can speak about their difference, without resorting to the precise time of language separation. To establish the proximity of two languages, one should compute the correlation coefficients of their phonological, morphological, syntactical, and other categories. It is the task of statistical language typology and dialectology. Grammatical statistics researches the quantitative relationships of morphological categories and parts of speech.

Syntactic statistics determines the frequency, distribution, and relationships of sen-tence structures and types of sentences. Semantic statistics describes the quantitative properties of semantic fields of different words.

Stylistic statistics studies quantitative features of style, correlation, and disputed authorship. The application of statistical methods to problems of disputed authorship has a rather long tradition in English with respect to the works of Shakespeare, Mark Twain, Hamilton, etc. It is often called stylometry. The study of the statistics of sound or word usage helps to catch the nuances of the text. The statistical spectra of sounds or words then and now may give a history of the development and changes of a language as it is and language habits of a particular writer. The use of stylometry in authorship attribution is based on the hypothesis that there are quantifiable features of style, characteristic to a particular writer. Authorship attribution is usually performed in the following way: is the literary opus under investigation more similar to the candidate author "A" or more similar to the candidate author "B" based on the accepted computed features? Parametric and nonparametric statistical tests are used to measure similarity, e.g. t-test (or Student's-test), chi-square test (or Pearson test), Mann–Whitney, Wilcoxon, and some other tests.

Typological statistics quantifies the so-called language universals. In the 1960s, Joseph Greenberg deduced 11 typological indexes to put languages in some order, e.g. index of synthesis, i.e. the ratio of the number of morphemes to the number of words, or index of complexity of morphological structure; index of agglutination, i.e. the ratio of the number of agglutinative inner morphological links and the number of words with inner morphological links; index of derivation, i.e. the ratio of the number of roots and the number of words; index of inflation, i.e. the ratio of the number of nonreflexive morphemes and the number of words, etc. These measures were further modified and complemented by different linguists to eliminate the shortcomings.

Some linguists include corpus linguistics, a branch of linguistics concerned with the study of language use by means of large text corpora into mathematical linguistics, because they use statistical methods. For the last 30 years, natural language processing has been developing rapidly. The work that has been done in this area has resulted in a large variety of systems constructed for specific applications. Very often, this particular field is called computational linguistics, and is considered to be a separate discipline in its own right. This strand of mathematical linguistics is close to machine translation and consists in using the computer as a tool in testing theoretical linguistic models like transformational grammars, Montague grammars, generalized grammars, etc. Computational linguistics include speech synthesis, speech recognition, constructing concordances, etc. It is advisable to include in mathematical linguistics as a discipline all areas of linguistics where various statistical counts and analyses are required.

References

Brainerd, Barron. 1974. Weighing evidence in language and literature: a statistical approach. Toronto: University of Toronto Press.

Burrows, J. F. 1987. Computation into criticism. Oxford: Clarendon Press.

Butler, Christopher. 1985. Statistics in linguistics. Oxford: Basil Blackwell.

Herdan, Gustav. 1966. The advanced theory of language as choice and chance. Berlin: Springer-Verlag.

Kenny, Anthony. 1986. A stylometric study of the New Testament. Oxford: Clarendon Press.

Ledger, Gerard R. 1989. Re-counting Plato. A computer analysis of Plato's style. Oxford: Clarendon Press.

Onishi, Masao. 1937. Articulation bases of ten major languages. A study from the view-point of occurrence of sounds. Study of Sounds 6 [in Japanese].

Oostdjik, Nelleke. 1991. Corpus linguistics and the automatic analysis of English. Amsterdam: Rodopi.

Piotrovsky, R. G., K. B. Bektaev, and A. A. Piotrovskaya. 1977. Mathematical linguistics [in Russian]. Moskva: Vys. Shkola.

Tambovtsev, Yuri. 1994. Dynamics of functioning of phonemes in sound chains of the languages of different structures [in Russian]. Novosibirsk: Novosibirsk State University.

———. 1994. Typology of orderliness of sound chains in language. Novosibirsk: Novosibirsk State University.

Teshitelova, Marie. Quantitative linguistics. Amsterdam: John Benjamins.

YURI TAMBOVTSEV

Maya and Mayan Languages

Sharer's *The Ancient Maya* (1994) begins with a clarification of the more standard uses of the terms Maya and Mayan, citing the first as both noun, as in the Maya, in reference to the Maya people, and as adjective, as in Maya writing. Mayan, on the other hand, is used with reference to the language family,

and the individual languages that make up that family, such as K'iche' Mayan, Yucatec Mayan, etc. These conventions will be adhered to in the present essay.

The Mayan language family consists of 28 extant languages and have approximately 6 million Mayan speakers covering a fairly contiguous geographical area encompassing southern Mexico, most of Guatemala, and parts of Belize, El Salvador, and Honduras. The number of speakers of individual Mayan languages ranges considerably, with speakers of K'iche', for example, spoken in southern Guatemala, numbering more than 1 million, while others, such as Itzá, spoken in the Peten of Guatemala are on the verge of extinction. The one group of Mayan speakers not geographically bound to the rest of the family, the Huastecs, migrated north some 4,000 years ago along the Gulf Coast to present-day Veracruz. It is believed that at approximately the same time, the Yucatecan branch of the Mayan family separated from the main body. This subgroup consists of four languages spoken throughout the Yucatan Peninsula, northern Guatemala, and parts of Belize. Yucatec Mayan, by far the largest of this subgroup in terms of speakers, numbering approximately 1 million, is the language of the three of the four surviving hieroglyphic books, or codices.

The main Mayan language body subsequently split into eastern and western subgroups, each with several offshoots, including, ultimately, the Cholan languages, the principal languages of Maya hieroglyphics. Much of recent inquiry and investigation into Mayan and among the Maya people has been generated by interest in the decipherment of the ancient Maya writing system, as explored thoroughly in Coe's *Breaking the Maya code* (1992).

In order to gain an appreciation of the richness of Mayan, it is important to define a few concepts and linguistic elements commonly addressed in the literature on this indigenous language family. Each will be defined and examples will be cited within the section in which they appear. The language examples cited herein, while limited to Yucatec Mayan, will serve to illustrate a number of features found cross-linguistically throughout the entire Mayan family. For more detailed explanations of a number of grammatical features in several Mayan languages, see Robertson's *The history of tense/aspect/mood/ voice in the Mayan verbal complex* (1992). The examples cited will utilize the standardized alphabet agreed to by *La Academia de la Lengua Maya de Yucatán, A.C., Mayáon, A.C.,* and others, in 1984, an important point as many recent publications and dictionaries, let alone older and colonial texts and materials, frequently do not. The topics presented in the following segments include the grammatical category aspect, the pronominal system, the concept of ergativity, the numerical system and numerical classifiers, and finally the phonological system.

Aspect

The term aspect is frequently overlooked in language teaching and learning. Comrie (1976) comments that in spite of the usage of more traditional terminology, for example, tense, with regard to the temporal relations expressed in sentences, such as *John painted the house last summer* and *John was painting the house last summer,* the distinction between these two is not one of tense, but rather of aspect. Both sentences occur in the past; thus; an additional explanation to account for the difference in meaning must be given. Aspect is the grammatical term used to describe how a situation or action is viewed. When the meaning expressed by the verb is viewed in its entirety as completed, the term perfective is used to indicate the aspectual category. When it is not viewed as completed, and makes no reference to the beginning or end of the action or state, the term imperfective is used.

Mayan languages do not have tenses, but denote temporal distinctions by means of a number of aspect markers and particles. Bricker et al. (1998) list the following aspect categories for Yucatec Mayan:

Assurative future	je'el…e'	Je'el in wilikech. I **will** see you.
Completive	t-…(-aj)	*Tin wilajech.* I saw you.
Compulsive	yaan	*Yaan in wilikech.* I must see you.
Definite future	k-…-j	*Kinj ilikech.* It is I who will see you.
Desiderative	taak	*Taak in wilikech.* I need to see you.
Durative	táan	*Táan in wilikech.* I am seeing you.
Habitual/Incomplete	k-…(-ik)	*Kin wilikech.* I see you.
Inceptive	jo'op'	*Jo'op' in wilikech.* I began seeing you.
Obligative	k'abéet	*K'abéet in wilikech.* I need to see you.
Proximate perfective	táant…e'	*Táant in wilikeche'.* I just saw you.

Additional aspect markers include constructs such as a continuous past (*laayli'…e'*), a remote future (*Bíin…*), etc. By means of these elements, Mayan speakers indicate their perceptions of the internal temporal components of a situation, without reference to external criteria needed for tense distinctions such as present, past, and future.

Pronominal System and Ergativity

The pronominal system of Mayan consists of three sets of pronouns, commonly referred to as Set A, Set B, and the Independent Set. Set A pronouns in Yucatec serve as subjects of transitive verbs, as subjects of intransitive verbs in the incompletive aspect, and as possessives. Set B pronouns serve as direct objects of transitive verbs and subjects of intransitive verbs in the completive aspect. The Independent pronouns are used for emphasis or clarification at the beginning of a sentence, with topical marker -*e'*, also at the beginning of a sentence or clause, and as indirect objects. Whereas the pronouns of Sets A and B are bound to an aspect particle or word, or to a verb, the Independent pronouns are free of such attachments, and hence the designation.

Set A pronouns

	Singular	Plural
1st person	in-	k-
2nd person	a-	a-...-e'ex
3rd person	u-	u-...o'ob

Set B pronouns

	Singular	Plural
1st person	-en	-o'on
2nd person	-ech	-e'ex
3rd person	-Ø	-o'ob

Independent pronouns

	Singular	Plural
1st person	teen	to'on
2nd person	teech	te'ex
3rd person	leti'	leti'o'ob

The uses of Set A and Set B pronouns summarized above illustrate a linguistic phenomenon common to perhaps one quarter of the world's languages, that of ergativity. Ergativity is the "term used to describe a grammatical pattern in which the subject of an intransitive clause is treated in the same way as the object of a transitive clause, and differently from transitive subject (Dixon, *Ergativity,* 1994:1)". In English, this could be rendered as *She see him* and *Him sees,* and *He sees it,* where *him* is the direct object of the transitive verb in the first example, and the subject of the intransitive in the second, complementary to the use of *He* as subject of the transitive in the third example.

The following figure shows a representation of Dixon's universal syntactic–semantic primitives, in which the letter A symbolizes the subject of transitive verbs, with S representing the subject of intransitives, and O, the object of transitives. In the nominative–accusative case marking pattern found in English, for example, the subject of transitive and intransitive verbs is treated equally, that is, they are the same; it is the object pronoun that is marked differently. In Mayan, an ergative–absolutive family of languages, the subject of intransitives and the object of transitives are marked in the same manner, and different from the subject of transitives.

$$\text{nominative} \left\{ \begin{array}{ll} A & \text{ergative} \\ S & \\ \end{array} \right.$$
$$\text{accusative} \quad\quad O \left. \begin{array}{l} \\ \end{array} \right\} \text{absolutive}$$
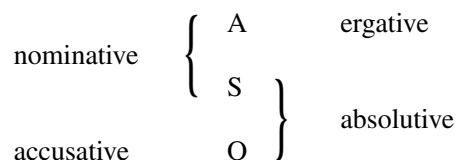
Many of the Mayan languages have developed a split ergative system when certain constructions follow the ergative–absolutive pattern of the ancestral Common Mayan language, and others follow a nominative–accusative pattern. This split in the usage of pronouns is aspectual in nature, with an incompletive aspect following the nominative–accusative configuration, and completive, the ergative–absolutive.

Numerical System and Numerical Classifiers

The Maya developed a vigesimal numbering system, that is, based on 20 rather than a decimal, base 10 system. Unlike English, the numbers in Maya cannot be used in isolation. For example, in response to the question *How many brothers do you have?*, you cannot say simply, *One.* The numbers are always used with a numerical classifier, denoting animacy, measure, or shape. A similar phenomenon in English can be seen with measurements and group classifications such as *a pinch of salt, a flock of geese,* and the like. A list of some of the more commonly used classifiers includes the following.

-cháach	a handful of something
-chi'náab	a measure from the thumb to the end of the forefinger
-kúuch	a load of something, typically carried on your back
-kúul	for plants
-lúub	a measure of distance of approximately four kilometers
-náab	measure from thumb to end of little finger; a handspan
-p'éel	for inanimate objects
-p'u'uk	a mouthful
-sáap	a measure of depth, equivalent to a fathom
-téen/-púul	number of times something is done
-ts'áan	a pair of things
-ts'íit	long, thin shape
-túul	for people, animals, etc.
-wáal	for flat, thin objects, such as tortillas
-xéet'	a piece of something

Phonology

Tozzer (1921), addressing issues of Mayan pronunciation from one of the earliest grammars written on Yucatec Mayan, that of Beltran de Santa Rosa dating back to 1746, says (translated):

> Because in this language you don't speak entirely like you write, nor do you pronounce many words according to the rules (and hence, some who speak it sound like foreigners or are judged as not speaking as they should; being thus, that they speak in accordance with the rules of the grammar) be advised that the use of synalepha and syncope is so necessary, that without hyperbole one may state, that all of the being and attractiveness of this language is the use of them and the main part of its grammar is their explanation.

> Syncope is no other thing than swallowing some word or some syllable, some vowel or consonant. And this figure so graces the Mayan tongue that without it it seems that her words become strange, unpleasant and in her cadence, ugly. In such a manner, that I can without recklessness say that almost one half of her words are syncoped or are syncopable.

Indeed, there are a number of phonetic changes in Mayan that clearly distinguish the written form from the spoken. Three such phenomena are syncope, the loss or suppression of one (or more) sounds in the interior of a word, synalepha, the fusion into one word from the final syllable of one word with the initial of the next, and finally, apocope, the cutting off or omitting of the last sound or syllable of a word. Occurrences of each can be seen in the following examples.

Syncope	• The transitive verb with a pronominal object may lose the **-i-** of **-ik** in the incompletive and the **-a-** of **-aj** in the completive: **Kin wilikech.** → **Kin wilkech.** • All polysyllabic transitive verbs lose the vowel of the temporal endings before the **-o'on** of the 1st person plural, **-e'ex** of the 2nd person plural and **-o'ob** of the 3rd person plural: **Tu ya'alajo'on.** → **Tu ya'aljo'on.** • In words of two syllables containing two similar vowels, the second is lost when the plural is used, a verbal pronoun is used, or a demonstrative pronoun is used: **xanab** → **xanbo'ob**
Synalepha	• Contraction of durative and A pronoun: **táan in** → **tin, táan a** → **tan** • Contraction of terminative and A pronoun: **ts'o'ok in** → **ts'in, ts'o'ok a** → **ts'a** • The negative **ma'** and the nominal pronoun: **ma' in** → **min**
Apocope	• Word final **-l** is lost in most cases: **Tun weenel** → **Tun weene.** • **j** → **Ø** optionally in many cases

An additional element of Mayan phonology of note is the predominance of stops among both consonants and vowels, a phenomenon whereby the flow of air through the oral cavity is briefly interrupted. An example of a glottal stop in English can sometimes be heard replacing the *t* sound in the word *button*, with the flow of air momentarily restricted at the vocal cords. All five of the vowels in Mayan may incorporate a glottal stop, as in *ma'*, no, and *si'*, firewood. All of the vowels may be reduplicated, in which the vowel is interrupted by the stop and then rearticulated, as in *ba'alo'ob*, things, and *tu'ux*, where.

The occurrence of stops on the consonants ch, p, and t, indicated in written form by the addition of an apostrophe, ch', p', t', is said to be phonemic in Mayan, as the articulation of these stops can serve to distinguish a word with them from another word without, as in *maak*, to cover, and *maak'*, to eat quickly, *chak*, red, and *ch'ak*, to cut with an ax.

A final distinctive phonological element common to many of the Mayan languages is that of pitch. Yucatec Mayan is a tonal language, with three distinctive tones. One is long and high, and the other is long and low, as in x*uul,* end, and x*úul,* planting stick, and *miis,* cat, and *míis,* broom. The third tone is realized on short vowels, and is considered neutral.

It is sincerely hoped that the preceding essay has provided you with a very brief but enticing introduction to the Maya and their languages, and that as a result of the planting of this infinitesimal grain of linguistic sand, your interest and enthusiasm for linguistics grows and develops into your own unique pearl.

References

Bricker, Victoria, Eleuterio Po'ot Yah, and Ofelia Dzul de Po'ot. 1998. *A Dictionary of the Maya language as spoken in Hocabá, Yucatán.* Salt Lake City, UT: University of Utah Press.

Coe, Michael D. 1996. *The Maya.* NewYork: Thames and Hudson, Inc., 5th edition, 1993.

Coe, Michael D. 1992 *Breaking the Maya code,* NY: Thames and Hudson, Inc.

Comrie, Bernard. 1976. *Aspect.* Cambridge: Cambridge University Press.

Dixon, R.M.W.1994. *Ergativity,* Cambridge: Cambridge University Press.

Hofling, Charles Andrew. 1991. *Itzá Maya texts with a grammatical overview.* Salt Lake City: University of Utah Press.

Landa, Diego de. 1975. *Relación de las cosas de Yucatán;* as *Landa's Relación de las cosas de Yucatán, a translation,* translated by Alfred Tozzer. Millwood, NY, Kraus Reprint Co.

Mayáon, A.C. 1984. *Ts'iibjuumt'aan u ti'al u ts'íibta'al maaya t'aan (Alphabet for Writing Yucatec Maya).* Felipe

Carrillo Puerto, Quintana Roo: Comité Ejecutivo de Quintana Roo.

Anonymous. 1996. *Popul vuh: the Mayan book of the dawn of life* translated by Dennis Tedlock. New York: Simon & Schuster.

Robertson, John S. 1992. *The history of tense/aspect/mood/voice in the Mayan verbal complex.* Austin: University of Texas Press.

Sharer, R.J. 1947. *The ancient Maya.* Stanford: Stanford University Press, 5th edition, 1994.

Suárez Molina, Victor. 1945. *El español que se habla en Yucatán: Apuntamientos Filológicos.* Mérida, Yucatán: Ediciones de la Universidad Autónoma de Yucatán, 3rd edition, 1996.

Tozzer, Alfred M. 1977. *A Maya grammar,* NY: Dover Publications, Inc.

MICHAEL VROOMAN

McCawley, James David

An enthusiastic student, James David McCawley was admitted to the University of Chicago in 1954 at age 16. He took his MS in mathematics four years later. He was awarded a Fulbright Fellowship in 1959–1960 to attend Westfälische Wilhelms University in Münster, Germany. Although the fellowship was to study mathematics, McCawley's interest began to be drawn more toward languages. He studied a variety of languages, including Dutch. After the Fulbright year, he returned to the University of Chicago, where he continued his exploration of languages, taking Japanese and falling in love with it. During this period, he read Noam Chomsky's seminal work *Syntactic structures*. Linguistics as a field of study beckoned. He applied to MIT and was accepted in their first Ph.D. class in Linguistics. In 1965, he completed his dissertation under the direction of Noam Chomsky: *The accentual system of modern standard Japanese*.

McCawley returned to the University of Chicago, where he received tenure in 1969 and became a full professor in 1970. In the summer of 1971, he met then Ph.D. candidate Noriko Akatsuka. After a whirlwind courtship, during which Dr. McCawley went to Japan to formally ask for her hand, they were married on Beethoven's birthday, December 16, 1971. Dr. Akatsuka joined Dr. McCawley at the University of Chicago.

Dr. McCawley's interests were broad-ranging. He published on elements of phonetics and phonology as well as on syntax and semantics. His early training in mathematics put him at ease with the abstract formal systems of syntactic form and semantic integration. One of his most influential works attempted to make these mysteries accessible to his colleagues and students: *Everything that linguists have always wanted to know about logic (but were ashamed to ask).* Chicago: University of Chicago Press, and Oxford: Blackwell. He was a dedicated and accessible teacher, sharing his library, notes, and insights. The courses he taught ranged from introductory linguistics courses, syntax and semantics, to 'tense and aspect', 'discourse structure', and 'speech acts' to 'tone in Bantu languages' and 'Japanese phonology'. His inventory was so extensive that rather than attempting a term-to-term rotation of classes offered, Dr. McCawley would circulate a list of topics and ask students to choose.

His stature as an internationally recognized scholar and his outstanding ability as a mentor won him a named chair at the University of Chicago. He became the Andrew MacLeish Distinguished Service Professor of Linguistics and of East Asian Languages and Civilizations.

He was widely sought after as a speaker and as a guest faculty member. He accepted visiting or temporary appointments in 17 Universities, including the University of Illinois, University of Gööteborg, Australian National University, University of Hawaii, University of New Mexico, Central Institute of English and Foreign Languages (Hyderabad, India), Guangzhou Institute of Foreign Languages, and National Tsing Hua University (Hsinchu, Taiwan).

He served the Linguistic Society of America in a number of positions: as a member of the Nominating Committee (January 1974 to December 1976; Chairman 1976); as a member of the Executive Committee (January 1978–December 1980); as Vice-President (1995); and as President (1996). He became a member of the American Academy of Arts and Sciences in 1983, and served a term as its President. Dr. McCawley was associate editor of five professional journals; consulting editor for four; and on the editorial board of another five.

Dr. James D. McCawley brought joy and enthusiasm to his study of linguistics. He regularly hosted parties to celebrate linguistics with students and colleagues. He

championed Hanggul Day, the day the Hanggul alphabet was officially adopted by Korean, as *the* linguistic holiday. He composed linguistic doggerel for the University of Chicago departmental holiday commemorating the movement of the offices from Goodspeed Hall, annually celebrated with songs and counterscholastic papers. James D. McCawley had a famous alter ego, Quang Phúc Đông, who produced fine studies in scatological linguistics. His students honored him with a festschrift in this little developed genre: *Studies out in left field: defamatory essays presented to James D. McCawley on the occasion of his 33rd or 34th birthday*.

In addition to linguistics and music, Dr. McCawley had a passion for good food. He was a gourmet and an excellent cook. A condition for guest lectures was often an eater's tour of the host city. He shared his knowledge of and pleasure in Chinese cuisine through his book: *The eater's guide to Chinese characters*.

Dr. McCawley was also a staunch Libertarian. He ran as the Libertarian candidate for the Trustee of University of Illinois in 1976, 1978, and 1980. Throughout his career, he refused to accept federal funding for his research or to review other scholar's proposals for such studies.

James D. McCawley is known for his pioneering work on the interface of syntax and semantics, for his skepticism of theory, and for his demand for rigorous treatment of data. His publication record is outstanding. A full list can be found on his web-page: http://humanities.uchicago.edu/depts/linguistics/faculty/mccawleycv.html He set the framework for studies of tonal and near-tonal systems in works such as 'Some tonal systems that come close to being pitch accent systems but don't quite make it' (1970) and 'What is a tone language?' (1978). He provided a clear critique of syntactic theory in works such as (1982) *Thirty million theories of grammar* and (1988) *The syntactic phenomena of English*. At the time of his death, Dr. McCawley had several works in progress, including two books. His students have prepared him another festschrift. His lasting contributions, both in print and through the impact of those he has trained in the field, form a modern linguistics respectful of data and meticulously formal.

Biography

James David McCawley was born James Quillan McCawley, Jr in 1938, in Glasgow, Scotland; post-World War II, he emigrated as a minor to the United States, acquired citizenship, and changed his name to James David McCawley. He earned an M.S. in mathematics from University of Chicago in 1958 and earned a Ph.D. in linguistics from Massachusetts Institute of Technology in 1965 for his thesis on 'The accentual system of modern standard Japanese'. He worked as a translator of Russain math books for University of Chicago 5–9/59, 8/60–1/61; for the mechanical translation group at MIT 1–7/62, 2–6/63; as a teaching assistant for German at MIT 9/62–1/63; and as a research linguist for IBM Watson Research Center, Yorktown Heights, NY, 6–9/64. He took up a position as Assistant Professor of linguistics, University of Chicago, 9/64 and held this position until 9/69 when he advanced to associate; then, in 9/70, he became a full professor of Linguistics. He was appointed the Currently Andrew MacLeish Distinguished Service; during his tenure at Chicago, he held visiting appointments at Tokyo Center for Advanced Studies in Language, 6–9/67; University of Illinois (Linguistic Institute), 6–8/68; Scandinavian Summer School of Linguistics, 8/69;University of Göteborg, 9–12/69; University of Michigan, 1–4/70; Fifth International Seminar on Theoretical Linguistics, Tokyo, 8/70; University of California at Santa Cruz, 6–8/71; Australian National University (Canberra), 6–8/73; University of Hawaii (Linguistic Institute), 7–8/77; University of Illinois at Chicago, 1–3/78; University of New Mexico (Linguistic Institute), 6–8/80; Central Institute of English and Foreign Languages (Hyderabad, India), 7–8/80; University of Maryland (Linguistic Institute), 7–8/82; Georgetown University (Linguistic Institute), 6–8/85; Linguistics, Institute on Transformational-Generative Grammar, University of Delhi, 8–9/85; Guangzhou Institute of Foreign Languages, 9/87; University of Arizona (Linguistic Institute), 6–8/89; and National Tsing Hua University (Hsinchu, Taiwan), 2–6/94. He served as editor for 15 journals, and was president, vice-president, and held other posts within the Linguistic Society of America; May 1983, he was appointed fellow of American Academy of Arts and Sciences. He was also three time Libertarian candidate for Trustee of the University of Illinois. In 1991, he received an honorary Ph.D. from University Göteborg, He authored 225 scholarly articles and seven books; he was also dedicatee of Zwicky, Salus, Binnick, and Vanek (eds.), Studies out in left field: defamatory essays presented to James D. McCawley on the occasion of his 33rd or 34th birthday (Edmonton: Linguistic Research, 1971; reprinted 1992, Amsterdam and Philadelphia: Benjamins), and of Brentari, Larson, and MacLeod (eds.), *The joy of grammar* (Amsterdam and Philadelphia: Benjamins, 1992); and one forthcoming. McCawley died on April 10, 1999.

References

Chomsky, Noam. 1957. Syntactic structures. The Hague: Mouton and Company.

McCawley, James D. 1970. Some tonal systems that come close to being pitch accent systems but don't quite make it. Papers from the 6th regional meeting, Chicago Linguistic Society, 526–32.

McCawley, James D. 1978. What is a tone language. Tone: a linguistic survey, ed. by V. Fromkin, 113–31. New York: Academic Press.

McCawley, James D. 1981. Everything that linguists have always wanted to know about logic (but were ashamed to ask). Chicago: University of Chicago Press and Oxford: Blackwell.

McCawley, James D. 1982. Thirty million theories of grammar. Chicago: University of Chicago Press.

McCawley, James D. 1984. The eater's guide to Chinese characters. Chicago: University of Chicago Press.

McCawley, James D. 1988. The syntactic phenomena of English. Chicago: University of Chicago Press.

Zwicky, Arnold, et al. 1971. Studies out in left field: defamatory essays presented to James D. McCawley on the occasion of his 33rd or 34th birthday. Edmonton, Alberta: Linguistic Research.

JUDITH M. MAXWELL

Meaning

The linguistic study of meaning has traditionally concentrated on investigating conceptual or logical meaning, known as 'sense'. However, logical or conceptual meaning is not the only kind of meaning that sentences or utterances convey. In terms of child development, the interpersonal meanings of language are primary. And utterances might tell you who the speaker/writer is (idiosyncraey), their age or when they were speaking (age), what activity they are engaged in (occupation), where they come from (provenance), their relationship with the hearer/reader (status), and whether they are speaking or writing or combinations of these, such as reading aloud (mode) (Crystal and Davy 1969). It is clear that all these are social and interpersonal meanings, except the last, which is textual.

Semantic and Pragmatic Meaning

The distinction between interpersonal and conceptual meaning is in partial correspondence with the distinction between pragmatic and semantic meaning. Pragmatics answers the question 'What did the speaker A mean by uttering sentence X to hearer B in context C?', whereas semantics answers the much simpler question 'What does sentence X mean?' In pragmatics, meaning is a matter of what effect the speaker intends to produce in the hearer by making an utterance, which makes it interpersonal. Semantics has to do with the meanings of sentences and how these meanings can be computed by the meanings of their component parts (clauses, phrases, words, and morphemes) (Leech 1983).

Of course, we naturally encounter and learn meanings in pragmatic and social contexts, especially in our first language. Investigation of semantic meaning always involves an idealization and generalization of the meanings of the language code, abstracted from particular contexts. Such an idealization was thought necessary in order to make any progress in studying linguistic meanings scientifically and logically. However, it is apparent that semantics on its own cannot give anything like a full meaning to a sentence; in other words, semantics underdetermines meaning. One most obvious problem area is *deictic terms* (from *deixis* 'pointing'), whose meaning depends on knowing who uttered them (to whom), when they were uttered, and where they were uttered. The pronouns *I* and *you*, and the adverbs *now, yesterday,* and *here* are clear examples. But deixis is much more widespread than its occurrence in a few isolated lexical items. Most tense markers or aspectual particles are time deictics. Past, present, future, and more complex tenses refer to periods of time relative to the time of utterance. 'I saw him here yesterday' is incomplete in meaning unless we know the time of utterance so that 'saw' and 'yesterday' refer to a period of time previous to that time. In addition, we have to know where the speaker is to understand 'here', and who the speaker is to understand the referent of 'I'.

A related aspect of the pragmatic/semantic distinction is the difference between reference and sense (Frege 1892). Reference is the relationship between a phrase and the thing in the world that it identifies. The sense, however, is the concept represented by the meaning of the phrase. The versatility of language lies partly in the fact that any one sense can be used to refer to many different entities. *Woman,* whose sense is, let us say 'adult female', might be used to refer to any one of the two billion women in the world. Just as the relationship between sense and referent is one-to-many, so is the

reverse relationship between referent and sense—in other words, there are many different phrases with different senses that could be used to refer to the same individual: *The former president of the United States, the man from Arkansas, this consummate orator, the sexual partner of Monica Lewinsky.*

While reference in normally conventional, in that a particular phrase guides us quite clearly to identify the referent, this is not the case with metaphor—where the reference is unconventional. And names have no or little sense—we simply have to learn the pairing between each individual person or thing and the language form that refers to them.

Coded and Inferred Meanings

Words or lexical items can be seen as signs making up the linguistic code. These signs consist of a signifier (phonological or graphological form) and signified (meaning), like two sides of a piece of paper (Ferdinand de Saussure 1960). In general, these pairings are thought to be arbitrary; otherwise, languages would resemble each other much more (there is nothing about the form *dog* that makes it more suitable than the French *chien* to act as a signifier of 'dog'). Hence, in terms of semiotics (the study of signs), most linguistic signs are symbols (involving no resemblance between signifier and signified) rather than icons (based on resemblance) (Peirce 1940:104). This arbitrariness is not absolute, however. Some signifiers are obviously iconic, based on ono-matopoeia, e.g. *splash*, and at a more subtle level non-English speakers can guess quite reliably which word *tin* or *thunder* is most suitable as a signifier for the signified 'thunder'.

However, linguistic communication in any language is also heavily reliant on non-coded meanings, created by pragmatic inference, and some languages are perhaps more reliant on inferences from context than others. Chinese, for example, without a highly developed system of tense for time reference, leaves it to be inferred from the context.

Meaning, Social Context, and Ideology

While some inferences and interpretations might depend upon supplying general knowledge, many meanings are dependent more specifically on awareness of the context of utterance. Apart from conversation and narrative, discourse types generally occur in specific institutional contexts—lectures in an academic institution, sales reports in a business context, golf score-cards at a golf course, and so on. The concepts of register and genre are an attempt to systematically link discourse types with their institutional contexts. Each context has certain meaning potentials—likelihoods that certain meanings will occur in them.

Ideology can be seen operating directly in the meanings of words and clauses. According to Benjamin Lee Whorf (1956) and his theory of linguistic relativity, the world does not come to our consciousness in ready-made categories, but different languages dissect the continuity of our experience in different ways. Whorf's hypothesis, unfashionable for many years due to the influence of universalism in linguistics, has lately been rehabilitated to some extent (Lee 1996; Gumperz and Levinson 1996). This tradition, taken over later by Critical Discourse Analysis, emphasizes that language and its meanings are not a transparent medium through which reality is observed, but distort and construct that reality, giving us the concepts through which we think. The languages we speak hand down to us ready-made categories, which we regard as commonsense. They thereby unconsciously carry with them an ontology or ideology of which we may not be aware.

The classic case is of kinship categories. In Thai, for example, the primary criterion for categorizing siblings is by seniority rather than by sex as it is in English; elder sibling /phî:/ and younger sibling /nó:ŋ/. And this difference in ontology has ideological implications. Senior siblings in Thai culture have greater rights and responsibilities than in most Western cultures, traditionally giving orders and making demands of younger siblings, but also taking care of them and their welfare. Furthermore, Thais use these words as terms of address and second person pronouns with friends and acquaintances as well as with blood siblings, so that in a friendly conversation with a new acquaintance, when it is not obvious from appearances which of the conversational partners is the older, at some point the question has to be asked about relative age, to facilitate polite discourse and correct interpersonal behavior.

More generally, we may just not take much notice of things if we have no words for them in our language. In Chinese, there is a common word for the groove that runs between the middle of the upper lip and the nose. There is no such word in English. Presumably, this contrast reflects the unimportance of this aspect of reality to English speakers, and its importance to folk physiognomy in China.

Sometimes, of course, ideology is more obviously encoded in words and then they may become the site of ideological struggle. Was the movement of white people into Australia or Israelis into Palestine 'settlement' or 'invasion'? These words and their meanings then become contested terms. There can often be attempts to redefine the meanings of terms as

in 'property is theft'. Alternatively, one could try to replace a fashionable term with another that better suits one's ideology. For example, economic *growth* might be replaced with *cancer*, since from an ecological standpoint, growth in mature economies actually harms the environment and threatens our life-support systems. In discourse, there is often an ideological struggle to get one's meanings and metaphors accepted as the conventional ones.

Structuralism and Deconstruction

A parallel emphasis on the determination of meaning by code has been found in studies of literary meaning. Structuralism emphasized that a speaker or writer does not impart meaning to utterances, but that the code, the linguistic system underlying them, produces it. This conviction led to literary analysis aimed at revealing the underlying structures, which, rather like the grammar or morphology of the linguistic code, determine the shape of the texts that realize them (see, for example, Vladimir Propp's (1958) *Morphology of the folk tale*).

Structuralism gave way to deconstruction under the influence of Roland Barthes and Jacques Derrida. Barthes (1970) emphasized the radical ambiguity of the sign as part of the code and the play of signifiers. For him, literary meaning could and should not be constrained within the concept of unity as structuralism tried to pretend. Barthes valued literature that was less than easy to read, less than *lisible*, because it breaks with the coded conventions. In this type of writable or *scriptible* literary text, the 'code' items are echoic, intertextual, fragments of voices from other texts and from other codes. Derrida (1967) with his notion of *differance* (difference/deferring), suggested that signifiers have shifting signifieds whose meaning is never complete, because the text at hand always has reverberating connotations and each use of a text further develops its meaning. More extremely, for him, language became a play of signifiers without any signifieds—with nothing existing outside the text, and no meaning or truth independent of language (Jefferson and Robey 1982).

Reader Response, Dialogism, and Intersubjective Meaning

Less extreme than Derrida's emphasis on radical instability and subjectivity of meaning are the theories of Stanley Fish (1980) and Mikhail Bakhtin and Valentin Volosinov (1973). For Fish, meaning is 'an action made upon a reader rather than a container from which a reader gets a message'. Meaning is the experience of the process of reading in the reader's

head. He illustrates this with the sentence 'That Judas perished by hanging himself there is no certainty', which begins with the presupposition 'Judas perished by hanging himself', leading us to anticipate that the sentence will end with the word *doubt*, but reversing our expectation of certainty by ending with the word 'certainty'! However, Fish showed that it is not the subjective meanings of the individual reader that constitute either what is literature or how it should be read, but rather the inter-subjectivity of the interpretive community, into which readers are inducted by learning the conventions of how to read and make meaning.

Bakhtin too, emphasized the intersubjectivity of meaning by insisting that it only exists in dialog. Meaning is like a spark between two electric terminals (the speakers), and hence 'word is a two-sided act. It is determined by whose word it is and for whom it is meant'. Intertextually, words bear the traces of their past uses by previous speakers, and language is therefore never in one's own words but other people's words: 'the trouble with words is that you don't know whose mouth they've been in'. These uses and traces tend to be evaluative and ideological, part of social struggle. It follows from this view that 'the analysis of meaning should be based not only, and even not as much, on linguistics as on metalinguistics which studies the word not within the system of language and not in a text which is removed from dialogical intercourse, but precisely within the sphere of dialogical discourse itself'. Bakhtin's theories raise radical questions about the distinction between analytic (definitional) and synthetic (descriptive) meaning and the tension between meaning stability and meaning change.

Meaning, Metaphor, and the Experiential Hypothesis

Another alternative to subjectivity can be found in the experiential hypothesis that grew out of studies of metaphor. In its infancy, during the heyday of Chomskyan linguistics, the science of semantics attempted to marginalize metaphor as an anomaly, because nonliteral meanings posed challenges to semantic rules like selectional restrictions—the rules that stipulate what semantic pairings were allowable between phrases in a syntactic relationship. Noam Chomsky held his famous sentence 'colourless green ideas sleep furiously' to be semantically ill-formed because *sleep* should be restricted to human or animal subjects, *green* is only allowed to premodify concrete, not abstract, things, and so on. However, it is possible for an inventive reader to interpret this sentence metaphorically: 'Although ideas about protecting the

environment are boring, ignoring and suppressing them will only delay an angry social reaction.'

Besides such challenges, George Lakoff and Mark Johnson (1980) made clear that metaphor could not be ignored as it was all-pervasive, even in the language of those who try to reject it. For example, John Locke in this passage uses the metaphors 'move', 'mislead', and 'cheat'.

> But yet, if we would speak of things as they are, we must allow that ... all the artificial and figurative application of words eloquence hath invented, are for nothing else but to insinuate wrong ideas, move the passions, and thereby mislead the judgment, and so indeed are perfect cheat. [*Essay concerning human understanding* Book 3, Chapter 10, p.105]

The claims of Lakoff's conceptual metaphor theory (Lakoff 1987) are that abstract thought is fully possible only through the use of metaphor, and that these conceptual metaphors originate in our bodily infant experiences. We have certain preconceptual experiences as infants, such as body movements, our ability to move objects, to perceive them as wholes and retain images of them; and certain image schemas that recur in our everyday bodily experience, e.g. containers, paths, balance, up and down, part and whole, and front and back. The hypothesis claims that most abstract concepts arise from these preconceptual physical experiences and schemas by metaphorical projection. For example, abstract concepts like amount are conceptualized by projection from the bodily experience of up and down, producing a number of lexicalized metaphors:

> The number of books printed each year keeps on **going up**.
> My income **rose/fell** last year.
> If you're hot, turn the heat **down**.
> He is **under**age.

Many of the basic links in conceptual metaphors can be traced back to metonymic relations such as cause and effect. Hence, for example, if we have more things in a pile it will be higher than when we have less things, a possible origin of this set of lexical items. And there is a well-established set of metaphorical vocabulary in English, and other languages, which conceptualizes anger as heat. The origin of this is quite obviously one of cause and effect—when we become angry, our bodies do actually rise in temperature.

In the light of the lexical evidence for this hypothesis, Lakoff went on to develop an Experientialist philosophy, rejecting the Objectivist and Subjectivist paradigms. The problem with Objectivism is the independence assumption, namely:

> Existence and fact are independent of belief, knowledge, perception, modes of understanding, and every other aspect of human cognitive capacities. No true fact can depend on people believing it, on their knowledge of it, on their conceptualization of it, or on any other aspect of cognition. (Lakoff 1987:164)

Lakoff cannot accept this assumption, because, although there is a reality out there, we do not have unmediated access to it, as our thinking is inescapably mediated by the metaphors we use. Even scientific theories and models are basically metaphorical hypotheses that can only approximate descriptions of reality, and a new scientific model will initiate a programe of research trying to establish what features of the model do or do not apply. On the other hand, Lakoff rejects subjectivism and relativism, because infant bodily experiences are universal; hence, our conceptual metaphors are held in common. Experientialism, then, steers a middle course between objectivist philosophical traditions, which assume that truth is something we can access independent of any description of it, and subjectivism, which believes that truth is a matter of individual belief, relative to circumstances. However, unlike Critical Discourse analysts, Lakoff sees the body, not culture and ideology, as the determinant of thought and ontology.

Objectivist views of meaning disregard the idea that ideology and meaning exist intersubjectively in a dialectical relationship with each other, or that our physical experience of embodiment determines thought. According to Lakoff, the objectivist hypothesis claims that reality and fact are independent of observation or perception, or, for that matter, language. Within the objectivist paradigm of semantics, there arose a truth-conditional approach to meaning: knowing the meaning of a sentence is equivalent to knowing the conditions under which that sentence would be true. Clearly, if language constructs reality rather than merely reflecting it, the truth-conditional theory of meaning becomes circular to some extent. This is related to another problem, which is that truth–conditional semantics has tended to restrict itself to analytical statements rather than synthetic statements. A synthetic statement is one that is contingently true or untrue, such as 'Beckham transferred to Real Madrid', whereas an analytical statement is one that is necessarily true, such as 'boys are male'. It is doubtful whether theories of meaning, which exclude synthetic statements, can be useful models for the meaning of natural languages in real–world contexts.

References

Bakhtin, Mikhail, and Valentin Volosinov. 1973. Marxism and the philosophy of language. Cambridge, MA: Harvard University Press.
Barthes, R. 1970. *S/Z*. Paris: Seuil.

Crystal, D., and D. Davy. 1969. Investigating English style Harlow: Longman.

Derrida, Jacques. 1967. La voix et la phenomene. Paris: Seuil.

Fish, Stanley. 1980. Is there a text in this class?. Cambridge: Cambridge University Press.

Frege, Gottlob. 1892. Uber Sinn und Bedeutung. Zeitschrift fur Philosophie und philosoph. Kritik 100 25–50.

Goatly, Andrew. 1997. The language of metaphors. London: Routledge.

Gumperz, John, and Steven C. Levinson. (eds.) 1996. Rethinking linguistic relativity. Cambridge: Cambridge University Press.

Halliday, M.A.K., and Ruqaiya Hasan. 1985. Language, context and text. Geelong: Deakin University Press.

Jefferson, Ann, and David Robey. (eds.) 1982. Modern literary theory. London: Batsford.

Johnson, Mark. 1987. The body in the mind. London: University of Chicago Press.

Lakoff, George. 1987. Women, fire and dangerous things. Chicago: University of Chicago Press.

Lakoff, George, and Mark Johnson. 1980. Metaphors we live by. Chicago: University of Chicago Press.

Lakoff, George, and Mark Turner. 1989. More than cool reason: a field guide to poetic metaphor. Chicago: University of Chicago Press.

Lee, Penny. 1996. The Whorf theory complex. Amsterdam: Benjamins.

Leech, G.N. 1983. Principles of pragmatics. Harlow: Longman.

Peirce, C.S. 1940. *The philosophy of Peirce:* selected writings, ed. by J. Buchler. London.

Propp, Vladimir. 1958. The morphology of the folk-tale. Bloomington: Indiana Research Centre in Anthropology.

Saussure, F. de. 1960. Course in general linguistics. London: Peter Owen.

Thomson, John B. 1984. Studies in the theory of ideology. London: Polity Press.

Whorf, B.L. 1956. Language thought and reality: selected writings of Benjamin Lee Whorf, ed. by John B. Carroll. Cambridge, MA: MIT Press.

ANDREW GOATLY

See also **Context; Ideology and Language; Metaphor**

Medicine and Language

The description of language in medical encounters is one of the oldest and most prominent topics in discourse analysis (the study of the sequences and organization of language in context). The topic is interesting for theoretical and applied reasons: Theoretically, the description of language in medical encounters contributes to our understanding of institutional interaction, symmetrical and asymmetrical roles and relationships as created and reflected by discourse, and specialized sequences within the interaction of medical encounters (Drew and Heritage). Practically, the description of language in medical encounters allows linguists to help medical professionals improve communication and to help patients and families work effectively with medical professionals.

In Western industrialized societies, the medical encounter has a recognizable event structure. One of the first articles on the medical encounter by British researchers Byrne and Long examined more than 2,000 medical encounters and identified the typical parts of an encounter: (I) relating to the patient (greetings and small talk); (II) discovering the reason for the encounter (the patient complaint); (III) conducting a verbal or physical examination, or both (the history and physical); (IV) consideration of the patient's condition (delivery of diagnostic opinion and information); (V) detailing treatment or further investigation (treatment and advice); and (VI) termination (small talk and closing). Each part of the encounter is associated with conventional discourse sequences: question and answer in the history and physical, for example, and imperatives in the delivery of treatment and advice. Byrne and Long found, however, that the discourse of the medical encounter is highly asymmetrical: it is the physician who interactionally controls most of the discourse. The physician asks the questions, controls the topics and their development, deflects or ignores patient topics or contributions that he or she deems irrelevant, provides the amount of medical information that he or she deems appropriate, and determines the amount of social talk in openings and closings. The institutional power and authority of the physician, as well as the relatively powerless institutional position of the patient, then, are created, reflected, and maintained by the asymmetrical discourse practices of the encounter.

The asymmetry of a medical encounter—the control of the physician over the discourse—was a topic of much investigation in early research on language in medicine. Some of these asymmetries were troubling because they were seemingly related to issues of gender, class, and ethnicity. For example, a number of studies described the ways physicians ignored topics mentioned by female patients, especially when those topics moved

away from the medical world and toward the patient's personal world with its complicated psychological and social conditions (Mishler; Waitzkin). Paget argued that sequences like the following show how female patients' topics can be ignored by physicians who are in control of the discourse:

(1) PT: (sigh) I dunno.
    I'm thinkin maybe it's a hormone deficiency or something.
    DR: Let me just look at the scar.
    No-no that's awright.

During the interaction of physicians and female patients, these kinds of asymmetries are found in topic control and development (Davis). Other studies noted the ways that medical care seems to be practiced differently when patients are from different social classes and ethnicities. Fisher studied encounters in two different gynecological clinics: a community clinic staffed by resident doctors-in-training and attended by mostly poor minority women and a faculty clinic staffed by university physicians and attended by mostly middle-class white women. She found that women in the community clinic received more treatment recommendations for hysterectomies than women in the faculty clinic and, in fact, were not told that options other than hysterectomy were available. In the decision-making sequences in the faculty clinic, physicians often used questioning strategies: for example, a question such as *Now did Karen explain to you the abnormal, what the abnormal Pap smear business is?* can function to give patients a chance to display their knowledge and competence as background for collaborative decision-making. Patients, too, in the faculty clinic used questioning strategies: for example, a question such as *Is that necessary?* opens up the discourse to the discussion of alternative treatment options. Physicians and patients in the community clinic, however, often did not use questioning strategies, thereby reducing the chances that the discourse of the encounter would develop into collaborative decision making. The discourse in the community clinic was more oriented to persuasional strategies justifying the physician's choice of treatment, as in the following sequences (note that there are no contributions from the patients in these sequences):

(2) DR: It (cancer) could come back…and for somebody your age, that's had your family, you're sure that you don't want children, I'd recommend a hysterectomy.
(3) DR: What you should do if you don't want any more children is have a hysterectomy. No more uterus, no more cancer, no more babies, no more birth control, and no more periods.

These persuasional sequences, with no mention of other treatment options, contrast with the presentational sequences used in the faculty clinic (note that the contribution from the patient significantly affects the course of the discourse):

(4) PT: Have a hysterectomy and that, I'm that, if there's an alternative. I'm terrified of operations.
    DR: Uh, okay, well, there certainly is an alternative, yeah, we can treat this by just freezing it here in the office and that usually will take care of it about 90% of the time.

Specific discourse practices such as questioning and the presentational or persuasional organization of information can thus have a direct effect on decision-making in the medical encounter. The troubling implication of this particular study was that the discourse of medical encounters seemed to differ on the basis of social class.

Much of this early research on the asymmetrical discourse of the medical encounter focused on question–answer sequences. Several classic studies showed how physicians control the discourse of the encounter by using questions that constrain patients to brief answers. One study looked at the discourse of ten medical encounters and found that patient-initiated questions (that is, utterances in question form that introduced new information or a new topic) comprised less than 1% of the discourse in a medical encounter (Frankel). Another study also found that physicians established question–answer discourse almost immediately after the beginning of the encounter: in this study, physicians typically interrupted the patient's account of the reason for the visit to begin asking questions after only 6 to 16 seconds (Beckman and Frankel). A third study defined questions more broadly as requests for information in any form but still found that patient questions comprised only 9% of the discourse in the 21 medical encounters that were analyzed (West).

These early studies of the language of the medical encounter were focused primarily on the ways physicians used discourse to maintain asymmetrical control of the discourse, with particular attention to the way that question–answer sequences turned the medical encounter into an interview (rather than a more symmetrical consultation or discussion). However, recent research in conversation analysis has been focused on the ways that both physicians *and* patients actively and collaboratively construct the interaction of the medical encounter as a type of institutional discourse (ten Have; Maynard). Sometimes, this construction achieves the asymmetrical discourse of the encounter as described above, with both physicians and patients working interactively to produce asymmetrical discourse. In a study of general practice encounters, for example, Heath found that patients rarely responded to

the delivery of a diagnostic opinion by a physician, not contributing an utterance at all after the delivery of diagnostic news, or offering only a minimal response:

(5) DR: Er, yes, (0.3) this one's blocked, the other one's not.
    PT: (1.2)
(6) DR: You've got, erm (0.8), bronchitis.
    PT: Er
    DR: (4.5) (doctor begins to write prescription)
        I'll give you antibiotics to take for a week.

In both sequences, the physicians offer the patients an opportunity to respond to the diagnostic news (note the pauses, with the number of seconds timed in parentheses), but the patients do not take the interactional opportunity to develop the topic. This kind of minimal response preserves the asymmetrical interactional and institutional power and role of the physician as the expert provider of diagnostic information. Patients, however, have their own interactional means of questioning or disputing a physician's diagnostic news. Although they typically remain silent or contribute only minimally to the discourse right after the delivery of diagnostic news, patients recycle the topic of diagnosis when they disagree with it. For example, a patient may repeat symptoms, emphasize their severity, or question the diagnosis during a different part of the encounter, often during the treatment recommendations or closing. In the following sequence, for example, the patient initially had not responded to the delivery of diagnostic news, but then produced an account of the severity of his symptoms after the physician had turned to the treatment activity of writing a prescription:

(7) PT: That I could understood (.) because it (.)
        it's the headaches was the thing that's
        got me, (0.4) More than anything else.
        (1.2) More than the devil in hell because
        they were getting more or less
        permanent yer know. (1.2) They were coming
        even when I was never—pain in the back of
        me neck.
        (28.00)
    DR: Right well I'll tell what we'll do Mister Tarrett.

Through their own interactional means, then, patients are able to question or challenge the physician's diagnostic opinion, even when they initially appear to have accepted it asymmetrically. A similar pattern was found in another study of the delivery of diagnostic news in a child assessment clinic: patients and families did not initially dispute the diagnostic opinion, but they found interactional ways to return to the topic if they disagreed with the diagnosis (Maynard).

Recent ethnographic research has looked at the ways that the discourse of physicians and patients reflects contextual dimensions such as power and expertise. One study looked at the ways that both physicians and patients made claims to power over the discourse of medical encounters (Ainsworth-Vaughn). In encounters in private practice between physicians and patients with long-standing relationships, the patients asked almost 40% of the questions, thereby controlling the discourse to a significant degree. An early study by Tannen and Wallat looked at the ways in which physicians construct frames for the delivery of medical information based on professional expertise. Two recent studies looked at the ways in which physicians acknowledge and utilize lay expertise in the discourse of medical encounters as well: in a children's disability clinic, families who display lay expertise about the condition of their children were able to move the discourse toward more collaborative decision-making, while families who do not display what physicians deem an appropriate lay expertise concerning their children's condition had medical encounters with a traditionally asymmetric discourse (Barton). These contextual dimensions are thus interactionally dynamic in their effect on the discourse of the medical encounter.

Analysis of the discourse of the medical encounter provides important information for the description of institutional interaction in general and physician–patient interaction in particular. This research has been used in a variety of applications. The curriculum of many medical schools now includes information about medical communication that draws on linguistic research. Community outreach programs, such as support groups, also draw on linguistic research to introduce families to ways of communicating effectively in the institutional context of medicine (e.g. asking questions about diagnostic labels, alternative treatments, and so on). Even consumer publications about medical communication have begun to draw on the research on medical discourse to advise patients and families in their efforts to work toward establishing collaborative discussion and decision-making in medical care.

References

Ainsworth-Vaughn, Nancy. 1998. Claiming power in doctor–patient talk. New York and Oxford: Oxford.

Barton, Ellen. 1999. The social work of diagnosis: evidence for judgments of competence and incompetence. Constructing (in)competence: disabling evaluations in clinical and social interaction, ed. by Dana Kavarsky, Judith Duchan and Madeline Maxwell. Mahwah, NJ: Lawrence Erlbaum.

Barton, Ellen. 2000. The interactional practices of referrals and accounts in medical discourse: expertise and compliance. Discourse Studies 2.

Beckman, Howard, and Richard Frankel. 1984. The effect of physician behavior on the collection of data. Annals of Internal Medicine 101.

Byrne, Patrick, and Barrie Long. 1976. Doctors talking to patients: a study of the verbal behaviours of doctors consulting in their surgeries. London: HMSO.

Davis, Kathy. 1988. Power under the microscope. Dordrecht, Holland: Foris.

Drew, Paul, and John Heritage (ed.) 1992. Talk at work: interaction in institutional settings. New York and Cambridge: Cambridge University Press.

Fisher, Sue. 1993. Doctor talk/patient talk: how treatment decisions are negotiated in doctor-patient communication. In Fisher and Todd.

Fisher, Sue, and Alexandra Todd (eds.) 1993. The social organization of doctor–patient communication, 2nd edition. Norwood, NJ: Ablex.

Frankel, Richard. 1979. Talking in interviews: a dispreference for patient-initiated questions in physician–patient encounters. Everyday language: studies in ethnomethodology, ed. by George Psathas. New York: Irvington.

ten Have, Paul. 1989. Talk and institution: a reconsideration of the 'asymmetry' of doctor–patient interaction. Talk and social structure: studies in ethnomethodology and conversation analysis, ed. by Deidre Boden and Don Zimmerman. Berkeley: University of California Press.

Heath, Christian. 1992. The delivery and reception of diagnosis in the general-practice consultation. In Drew and Heritage.

Maynard, Douglas. 1991. Interaction and asymmetry in clinical discourse. American Journal of Sociology 97.

Maynard, Douglas. 2002. Bad news, good news: conversational order in everyday talk and clinical settings. Chicago: University of Chicago Press.

Mishler, Elliot. 1984. The discourse of medicine. Norwood, NJ: Ablex.

Paget, Marianne. 1993. On the work of talk: studies in misunderstandings. In Fisher and Todd.

Silverman, David. 1987. Communication and medical practice. London and Thousand Oaks, CA: Sage.

Tannen, Deborah, and Cynthia Wallat. 1993. Doctor/mother/child communication: linguistic analysis of a pediatric interaction. In Fisher and Todd.

Waitzkin, Howard. 1991. The politics of medical encounters: how patients and doctors deal with social problems. New Haven, CT: Yale University Press.

West, Candace. 1984. Routine complications: troubles with talk between doctors and patients. Bloomington: Indiana University Press.

ELLEN BARTON

Meillet, Antoine

Antoine Meillet's intellectual and scientific life was organized around two fundamental axes: comparative linguistics and general linguistics. Up until 1906, his publications showed his interest in philology in the lineage of Ferdinand de Saussure, as well as his wish to describe and analyze culture in the semantic perspective of Michel Bréal. After 1906, Antoine Meillet became more familiar with the linguistic theories of the French sociologist Emile Durkheim (1858–1917). Then, he widened the fields of investigation and took an increasing interest in the definition of the general laws of linguistics.

Antoine Meillet gave a new vigor to comparative linguistics. For him, the analysis of the grammar of a language went hand in hand with the study of the history of this language. In his historical studies, he was concerned with achieving the highest degree of accuracy possible. He always maintained that linguistic forms cannot and must not be analyzed individually, but the whole encompassing systems have to be taken into account. Therefore, the exhaustive description of given languages was absolutely indispensable. These ideas are clearly formulated in his *Traité de grammaire comparée des langues classiques* (Essay on comparative grammar of classical languages). In the French edition, more than 700 pages refer not only to the linguistic relationship between Greek and Latin but also detail the development of both the Greek and the Latin linguistic system.

Ferdinand de Saussure's work *Mémoire sur le système primitif des voyelles dans les langues indo-européennes* (Report on the primitive system of vowels in Indo-European languages) (1878) made such an impression on Meillet that he used this work as a basic source and often quoted from it. He borrowed the notion of "system" from Saussure and it became the foundation of his work. The *Introduction à l'étude comparée des langues indo-européennes* (Introduction to the comparative study of the Indo-European languages), which explains and develops relationships between various Indo-European languages, is one of the most brillant illustrations of the Saussurean approach.

He also developed a theory of linguistic differentiation in order to explain how linguistic changes that occur at the point of origin of a given language family have less impact on languages in areas that are geographically remote from the point of origin. These more remote languages are thus more likely to retain archaic characteristics from their ancestor language. The impressive number of his contributions to the study of various languages (Armenian, Celtic, German,

Greek, Iranian, Latin, Polish, Slavic, etc.) shows the fruitfulness of his method and his way of thinking. His analyses were of such high quality that they still remain a model for research.

Antoine Meillet's theories on meaning have also remained influential to this day. During the first period of semantic studies, roughly from 1883 to 1931, Darwinian evolutionism gave rise to two opposing arguments. In France, Arsène Darmesteter was a defender of the organicist thesis, which maintained that languages and words must be considered as living organisms, similar to plants or animals. Thus, he argued that while the laws governing meaning may be described as fixed and timeless entities, words are born, live, change, and die according to natural selection. In *Essai de Sémantique* (Essay on semantics), published in 1897, Michel Bréal took up the opposite point of view, saying that language has no reality outside of human activity. For Bréal, the semantic laws are psychological or intellectual laws and govern language-specific patterns of words and sentences. This approach resonated with Antoine Meillet, for whom semantics always had a historical dimension. In "Comment les mots changent de sens" ("How words change meaning"), part of *Linguistique historique et linguistique générale* (Historic linguistics and general linguistics), he argued that modifications in the meaning of words depend on three interdependent sets of factors: linguistic ones (language structures), historical ones (cultural contacts), and social ones.

Throughout his career, he increasingly focused on the idea that language and the social environment are closely intertwined. In his studies on the period of bilingualism in France at the time of the Germanic invasion, he pointed out that antagonistic forces exist among communities, which must be taken into consideration in the description of the linguistic landscape. Due to these efforts, he was considered by some as the founder of a French sociological school. While this may not be completely accurate, his ideas certainly revolved around language as a social fact and he worked on topics that even today remain partly uncharted territory.

While Antoine Meillet is especially known as a comparative and general linguist, his book *Les Langues dans l'Europe nouvelle* (The tongues in New Europe), 1918, and *Les Langues du monde* (The tongues of the world), coedited with Marcel Cohen in 1924, also showed his great interest in dialect study and typology.

Famous for his publications and his personality, Antoine Meillet became a leader of a linguistic school. Many people were influenced by his theories and his works, like Émile Benveniste, Marcel Cohen, Gustave Guillaume, Louis Hjelmslev or André Martinet, and even Leonard Bloomfield in the United States admitted that his research on sentence structure were inspired by Meillet's ideas. Since Antoine Meillet's work was both innovative in its theoretical approach and rigorous in its implementation, and since it exhibited an interest in all languages, it can still be used as a stepping stone for further linguistic research.

Biography

Antoine Meillet was born in Moulins, France on November 11, 1866. He graduated at the Sorbonne in Paris, in 1885. In 1890, he conducted one year of fieldwork in the Caucasus, studying modern Armenian. Then, he became director of comparative Indo-European studies at the School of Advanced Studies in Paris. He received his Ph.D. in 1897 for his works on Old-Slavonic in *Recherches sur l'emploi du génétif-accusatif en vieux-slave* (*Researches on the Use of genitive-objective Case in Old-Slavonic*), his first thesis, and on the Indo-European root *men* in *De indo-europaea radice *men "mente agitare"* (*About the Indo-European Root *men*), his supplementary thesis in Latin. He was Professor of Armenian at the School of Advanced Studies in Paris in 1902. At the end of World War I, he contributed as an expert to the definition of linguistics groups, languages, and political boundaries. Then, he became secretary of the Société Linguistique de Paris (Paris Linguistics Society), before being elected to the Académie des Inscriptions et Belles-Lettres (French Academy of Inscriptions and Letters) in 1924. He retired in 1932. Meillet died in Châteaumeillant, France on September 21, 1936.

References

Auroux, Sylvain (ed.) 1988. Antoine Meillet et la linguistique de son temps. Histoire-Epistémologie-Langage X, (2). 5–348

Meillet, Antoine. [1903] 1936. Esquisse d'une grammaire comparée de l'arménien classique (Outline of a comparative grammar of classical Armenian), Vienne: Impr. des PP. Mêkharistes.

——— [1903] 1978. Introduction à l'étude comparative des langues indo-européennes (Introduction to the comparative study of the Indo-European languages), Alabama: University of Alabama Press.

——— [1913] 1975. Aperçu d'une histoire de la langue grecque (Outline of a history of the Greek language), Paris: Klincksieck.

——— [1915] 1931. Grammaire du vieux perse (Old-Iranian grammar), Paris: Champion.

——— [1921] 1982. Linguistique historique et linguistique générale (Historic linguistics and general linguistics), Paris: Champion and Genève: Slatkine Reprints.

——— [1925] 1979. Traité de grammaire comparée des langues classiques (Essay on comparative grammar of classical languages), Paris: Champion.

——— [1928] 1984. La méthode comparative en linguistique historique (The comparative method in historic linguistics), Paris: Champion.

——— [1936] 1951. Linguistique historique et linguistique générale II (Historic linguistics and general linguistics II), Paris: Klincksieck.

Swiggers, Pierre. 1996. Lexicon grammaticorum, 622–4 Tübingen : Max Niemeyer.

THIERRY PONCHON

See also **Bloomfield, Leonard; Martinet, Andrè; Saussure, Ferdinand de**

Metaphor

Metaphor and metonymy are two closely related concepts. They have been treated for centuries as types of figures of speech. Therefore, their study was thought to be the business of literary scholars, rather than of grammarians or linguists. Metaphor is defined from this 'classical' perspective as a figure of speech in which one word is used to indicate something different from the literal meaning, so that one thing or idea is likened to a different thing or idea. In a sentence like *He had a heart of stone*, the word *stone* is not used in its normal physical sense; rather, it conveys the idea of insensitivity, by likening this moral attitude to the hardness of stones. The classical view confines metaphor to ornate, flowery language, and assumes that the interpretation of any metaphorical use of language can be wholly reduced to literal meaning, which is its 'real' meaning (i.e. the 'real' meaning of the phrase *of stone* in this context is 'insensitive'). Conventional, automatic metaphorical expressions like *The* foot *of the mountain* are thus regarded as established literal meanings of words and cease to be metaphorical—they are called 'dead' metaphors.

Some Precognitive Modern Theories of Metaphor

The contemporary theory of metaphor that has had the strongest impact on linguistics is the cognitive theory of metaphor (see below). Among precognitive modern theories of metaphor, we can single out the 'interaction theory' and the 'pragmatist' position, which have also been followed by numerous linguists. The interaction theory, due to Max Black, claims that metaphor is the result of a semantic interaction or tension between a metaphorical term and its context. The tension between *stone* and *heart of* leads to construing the former as 'insensitivity', its 'real' meaning. The pragmatist position, due to Paul Grice and John Searle, holds that metaphor has to be reduced to the 'literal' meaning by applying a series of pragmatic principles of conversation, like conversational maxims. Some important papers representing these and other modern theories can be found in Ortony (1993).

The Cognitive Theory of Metaphor

This theory has revolutionized the scientific study of metaphor and metonymy, has dominated it for the past 20 years, and has attracted large numbers of linguists, psychologists, and other kinds of scientists to this field of research. (For the distinction and connections between metaphor and metonymy, see *Metonymy*.) The theory was first proposed in Lakoff and Johnson (1980). On empirical grounds, its authors criticized the basic assumptions of the traditional theory of metaphor and of most of the modern ones, especially the assumption that the meaning of metaphorical expressions can be reduced to 'literal' meaning, their 'real' meaning, and that metaphors had no serious conceptual import. The Cognitive Theory of Metaphor (CTM) evolved as an essential part of the approach known as Cognitive Linguistics (see *People, George Lakoff; Ronald Langacker*); a brief introduction is Ungerer and Schmid (1996). Let us now review some of the main tenets of the CTM:

(1) Metaphor makes it possible for people to understand new, complex, or abstract areas of experience ('domains') on the basis of more accessible domains, which are understood in their own terms; for instance, the abstract domain of life is often understood as a journey along a path (e.g. *He's at a crossroads* [in life]). The most basic metaphors have as input or 'source' domains universal physical notions like 'verticality', 'container', 'path', etc., known as 'image schemas' (Johnson 1987).

(2) Therefore, metaphor is not just a matter of language use but also, and fundamentally, a matter of thought, of conceptualization. Creative, conscious, unconventional metaphors are claimed to be usually extensions of automatic, unconscious, conventional metaphors.

(3) It is therefore important to distinguish between *conceptual metaphors* and *metaphorical expressions*. A given conceptual metaphor may be expressed by all kinds of linguistic and non-linguistic means (words, sentences, gestures, etc.); some of these linguistic expressions may cease to be used metaphorically but the conceptual metaphor may still motivate other expressions. For instance, in Old English *sad* originally meant 'full of food' and metaphorically 'full of sensations', the indirect source of its present sense 'sorrowful'. *Sad* stopped being a metaphorical expression when its original physical sense was lost. However, the conceptual metaphor PEOPLE ARE CONTAINERS FOR SENSATIONS, which indirectly motivated its metaphorical senses, is alive in countless other expressions (full *of joy*, *I am* fed up *with his behavior, sorrow*ful, etc.).

(4) Metaphors consist of fixed multiple simultaneous projections (or 'mappings') from the so-called 'source' domain onto the so-called 'target' domain. A sentence like *I see what you mean* is an expression of a conceptual metaphor that maps the conceptual structure of SEEING onto that of UNDERSTANDING, and which makes immediately comprehensible, without necessarily involving any pragmatic rules of interpretation, scores of other conventional expressions of the same metaphor like *Your arguments are* transparent, or *Your theory has thrown* light *on this problem*. (This is evidence against pragmatist theories.) Metaphorical mappings are fairly *systematic*, but they are only *partial* (about this, see Lakoff 1993).

(5) The mapping is always *unidirectional*: only the source is projected onto the target domain. Therefore, simultaneous *bidirectional* metaphorical projections do not exist, as the interaction theory claims. In the conceptual metaphor PEOPLE AS ANIMALS (as in *Don't* snap *at me* / *Their love* nest *has been discovered*), we project an aspect (aggressive behavior, living place) of some animals (dogs or birds) onto some aspects of people (anger, meeting point), but no aspect of people is mapped onto animals by virtue of *this* metaphor.

This would be done by a different metaphor, ANIMALS AS PEOPLE, as in *It's a* noble *dog*.

A very recent development within cognitive linguistics is the theory of 'blending' (due to Gilles Fauconnier and Mark Turner), which regards metaphor and metonymy as just two manifestations of a general mapping ability.

The CTM has been applied to the study of all aspects of language (lexicon, especially polysemy, grammar, discourse, and conversation) and to many other disciplines, from artificial intelligence to the study of literature, philosophy, anthropology, law, religion, etc.

References

Barcelona, Antonio (ed.) 2000. Metaphor and metonymy at the crossroads. A cognitive perspective. Berlin and New York: Mouton de Gruyter.

Cameron, Lynne, and Graham Low (ed.) 1999. Researching and applying metaphor, Cambridge and New York: Cambridge University Press.

Gibbs Jr., Raymond W.1994. The poetics of mind. Figurative thought, language and understanding. Cambridge and New York: Cambridge University Press.

——— and Gerard Steen (eds.) 1999. Metaphor in cognitive linguistics, Amsterdam and Philadelphia: John Benjamins.

Johnson, Mark. 1987. The body in the mind: the bodily basis of meaning, imagination, and reason. Chicago and London: The University of Chicago Press.

Kövecses, Zoltán. 1990. Emotion concepts. New York: Springer-Verlag.

Lakoff, George. 1987. Women, fire and dangerous things. What categories reveal about the mind. Chicago and London: The University of Chicago Press.

Lakoff, George. 1993. The contemporary theory of metaphor. In Ortony, pp. 202–51.

Lakoff, George, and Mark Johnson. 1980. Metaphors we live by. Chicago and London: The University of Chicago Press.

Lakoff, George, and Mark Turner. 1989. More than cool reason: a field guide to poetic metaphor. Chicago and London: The University of Chicago Press.

Ortony, Andrew, (ed.) 1993. Metaphor and thought. Cambridge and New York: Cambridge University Press.

Ungerer, Friedrich, and Hans-Jörg Schmid. 1996. An introduction to cognitive linguistics. London and New York: Longman.

ANTONIO BARCELONA

Metathesis

'Metathesis' refers to several related processes in which the linear order of two elements is reversed, as when *ask* is pronounced *aks* in dialectal English. Although in the broadest sense metathesis is the transposition of any two linguistic units, such as syllables or words (*She will go —> Will she go?*), it most commonly denotes the transposition of two sounds: consonant with consonant, consonant with

vowel, or vowel with vowel. In some cases, not two whole sounds, but just a feature of the sounds, for example vowel length, is transposed: *basile:os* 'of a king' in Ancient Attic Greek corresponds to *basileo:s* in Ancient Ionic Greek (the colon indicates a lengthened vowel). A thorough description of metathesis must begin with the following four parameters:

(1) Synchronic ~ diachronic. Synchronic metathesis occurs within one chronological period. Diachronic metathesis takes place from one time period to another, e.g. from Middle English to Modern English.

(2) Adjacent ~ nonadjacent. Adjacent metathesis occurs when two contiguous sounds are transposed. With nonadjacent or 'long-distance' metathesis, the transposed sounds may be separated by one or more intervening sounds.

(3) Regular ~ sporadic. Regular metathesis applies consistently, to many different words. Sporadic metathesis is restricted to only a few words and occurs haphazardly.

(4) Abrupt ~ gradual. Abrupt metathesis completely transposes sounds in a single step. Gradual metathesis can best be understood as a type of chain reaction, a series of sound changes over time that eventually inverts two sounds. Gradual metathesis therefore uses one or more intermediate stages between the original order and the final transposed order.

A closer look at metathesis reveals four major types of inversion, pairing regular or sporadic with synchronic or diachronic, each one a unique combination of the above parameters with other linguistic causes and tendencies.

One type of synchronic metathesis is an abrupt regular process that applies to adjacent sounds in order to mark certain grammatical relationships between words. For example, in Rotuman, a Malayo-Polynesian language of the central Pacific, metathesis marks the distinction between definite and indefinite: /hosa/ 'the flower' vs. /hoas/ 'some (unspecified) flower'. Other languages where metathesis has a grammatical function include Clallam, Leti, Cherokee, Zoque, and Kwara'ae.

Regular synchronic metathesis is generally the product of a previous diachronic metathesis. Speaker variation between metathesized and unmetathesized forms, initially nothing more than variation in the pronunciation of a word during an earlier chronological stage of the language, is eventually sorted out over time by assigning each of the two variants to one of two options within a specific grammatical category, e.g. singular vs. plural or definite vs. indefinite.

In the absence of grammatical conditioning, synchronic metathesis is sporadic and abrupt, motivated by speech errors ('slips of the tongue') or by the influence of words related in form or meaning. For example, the English cases of *irrelevant ~ irrevelant* and *integral ~ intregal/intrical* may be due to the influence of *revelation* and *intricate*, respectively. As these examples show, sporadic cases of synchronic metathesis often occur when the inverted sounds are not adjacent to one another, and they may involve the transposition of two sounds or the relocation of one sound into another syllable.

If sporadic synchronic metathesis occurs in certain words with sufficient frequency among speakers in a given speech community, it may become a permanent diachronic sound change over time. Sporadic diachronic metathesis includes: (1) the long-distance (reciprocal) metathesis of two sounds, often *l* or *r:* e.g. Latin *parabolam* 'word' developed into Spanish *palabra;* (2) the transposition of one sound into a neighboring syllable: Latin *capra* 'goat' became dialectal Italian *crapa*; and (3) the inversion of two adjacent sounds, as with Old English *bridd* and Modern English *bird.*

A sporadic diachronic metathesis can become regular (permanently invert the same two sounds in many different words) if it fulfills a specific structural purpose. For example, a preference for syllables ending in a vowel may have caused the change from Proto-Slavic **melko* to South and West Slavic *mleko* 'milk' (the asterisk indicates that the relevant word form is not actually attested, but reconstructed on the basis of modern words). Similarly, less favored consonant sequences such as /dl/ and /nr/ are often metathesized in order to locate the more sonorant (vowel-like) sound in the syllable-final position and the less sonorant sound at the beginning of the next syllable, e.g. Latin *titulum* 'title' most likely developed first into **tidle*, which then metathesized into Old Spanish *tilde;* Latin *generum* 'brother-in-law' became **yen.ro* and then Spanish *yer.no*. In contrast to these examples, a sporadic metathesis without structural motivation, as in the *irrelevant ~ irrevelant* example*,* cannot become regular, i.e. this same process will not spread to other words with *l* and *v* in similar positions.

Regular diachronic metathesis is often the result of a series of interrelated sound changes over time, a gradual chain reaction of regular changes whose last stage creates a transposed version of the original sequence. For example, French *Breton* became **brtõ*, which then developed into Le Havre French *bertõ*. This is an example of 'pseudometathesis,' since the

two sounds *e* and *r* are not transposed directly. Rather, the *e* was first lost and then an *e* was inserted into a different position. If only the original word and the final result were compared, the modern word would seem to be the product of metathesis, when in fact it results from several nonmetathetical processes.

A gradual diachronic transposition is often the case with consonant + vowel sequences, where *i* or *u* reduces to *j* or *w*, respectively. These sounds can then merge with adjacent consonants—they become 'secondary articulations' of these consonants—and later give rise to vowel developments on the other side of the consonant. Examples of such chain processes are as follows: Ancient Greek *phéresi* > *phéresʲ* > *phérejs* 'you carry'; Latin *sapui* > **sabwi* > **sabʷi* > **sawbe* > **sowbe* > Portuguese *soube* 'I knew'. The vowels thus gradually wandered 'through' a consonant. The result, again, resembles the product of metathesis. Consonants may also reduce to secondary articulations before reappearing on the opposite side of a vowel. For example, an *h* may temporarily become a breathy-voice articulation on the neighboring vowel and then resurface as a full consonant on the other side: Cayuga /akekaha̰/ >

*agéka̰a̰ > [agékhaa̰] 'my eye'. As before, the *h* actually moved through the vowel *a* in several steps, but the result resembles a one-step metathesis.

References

Alexander, James D. 1985. R-Metathesis in English: a diachronic account, Journal of English Linguistics 18. 33–40

Besnier, Niko. 1987. An autosegmental approach to metathesis in Rotuman. Lingua 73. 201–23

Blevins, Juliette, and Andrew Garrett. 1998. The origins of consonant–vowel metathesis. Language 74(3). 508–55

Grammont, Maurice. 1933. Traité de phonétique. Paris: Delagrave; 8th edition, 1965.

Hock, Hans Henrich. 1985. Regular metathesis. Linguistics 23. 529–46

Hogg, Richard M. 1977. Old English *r*-metathesis and generative phonology. Journal of Linguistics 13. 165–75

Hume, Elizabeth. 1988. Metathesis in phonological theory: the case of Leti. Lingua 104. 147–86

Ultan, Russell. 1978. A typological view of metathesis. Universals of human language, Vol. 1, ed. by Joseph H. Greenberg. Stanford: Stanford University Press.

Vennemann, Theo. 1988. Preference laws for syllable structure and the explanation of sound change. Berlin: Mouton de Gruyter.

KENNETH J. WIREBACK

Metonymy

Metaphor and metonymy are two closely related concepts. Both have been treated for centuries as figures of speech, and are as such regarded to be the business of rhetoric or literary studies rather than linguistics. The traditional theory of metonymy defines it as the 'use of the name of one thing for that of another associated with or suggested by it'. (*Webster's new twentieth century dictionary*, 2nd edition.) In *The White House issued a statement*, the phrase *the White House* is used instead of the phrase *the President*. This traditional theory treats metonymy as a matter of language use; its conceptual nature is overlooked. The meaning of metonymic expressions is also assumed to be wholly translatable into literal meaning. Metonymic expressions are considered to be necessarily referential (in the technical sense of designating an extra-linguistic entity). Most twentieth-century treatments of metonymy have essentially maintained the same views; a good illustration is Ullmann (1962).

The Cognitive Theory of Metonymy

This theory emerged alongside the cognitive theory of metaphor, but the nature of metonymy and its relevance for thought and language have only recently begun to be investigated in depth. In the standard form of the Cognitive Theory of Metonymy (CTMY), as in Lakoff and Turner (1989), metonymy is defined as a mapping with a primarily referential purpose, in which the source and the target are conceptual entities in the same domain. 'Mapping' means 'conceptual projection'. In the previous example, the concept of the White House (the 'source') is mapped onto the concept of the President (the 'target'): the President is primarily conceptualized as someone who lives in that building. This allows the phrase *the White House* to *refer to*, i.e. to designate, the person in office as the President of the United States at a given time. Finally, a place and the people or things located in it are so closely linked in experience that they can be said to be in the same 'domain' or area of experience.

Metonymy can link a part to a whole, as in *We need a couple of extra* hands *in our factory* (HANDS [BODY PART] FOR PEOPLE), a whole to a part, as in The Times *is here* (NEWSPAPER COMPANY FOR REPORTER), or a part to a part, as in *He has a good* pen (PEN [INSTRUMENT] FOR ACTION [WRITING]). This tripartite classification continues the traditional theory, except that the latter reserves the term *synecdoche* for PART-FOR-WHOLE metonymies.

In the CTMy, metonymy is a primarily conceptual phenomenon, rather than just a matter of language use, so that conceptual metonymies have to be distinguished from metonymic expressions; metonymy is systematic, as it responds to general patterns like WHOLE FOR PART or to more specific patterns like INSTRUMENT FOR ACTION. And the meaning of metonymic expressions cannot be simply reduced to literal meaning.

Some of the main areas of current debate in the CTMy are the following:

(1) Referential function. Some cognitive linguists still hold the traditional view that every metonymic noun phrase must be referential. But most cognitive linguists note that, even though referring is a very frequent function of metonymy, metonymic noun phrases are often also used nonreferentially. For instance, in *John is a good* hand *at fencing*, the predicate noun phrase does not designate an individual; it just mentions a property.

(2) The nature of the relationship between source and target. Some linguists avoid treating metonymy as a mapping process, because, unlike metaphor, metonymy apparently does not consist of systematic multiple mappings. An approach that has recently won widespread support is that of regarding the metonymic source as a 'cognitive reference point', which, in an appropriate context, can cause the mental activation of the target (see Kövecses and Radden 1998; Langacker 1999): In *The White House issued a statement*, a LOCATION causes the mental activation of the PERSON LOCATED. It also imposes a certain perspective on that person. As we saw earlier, this imposition is a type of mapping .

(3) The distinction from, and interaction with, metaphor. The key property distinguishing them, according to the standard view in the CTMy, is that metaphor is a mapping across *different* domains, whereas metonymy is a mapping within the *same* domain. Consider the metaphor LIFE IS A JOURNEY, as in *I have gone a long way* (said about one's life): life and journeys are in two different domains. In the metonymy CONTAINER FOR CONTENT, as in *He drank a* glass, the source (the glass) and the target (the contents of the glass) are in the same domain, namely, the domain of 'containment'. This criterion, however, does not always work clearly. For instance, the source and the target of PEOPLE ARE ANIMALS (*Richard is a lion*) are in the same domain, the domain of LIVING BEINGS. This and other cases have led some linguists to claim that an absolute distinction between metaphor and metonymy cannot be maintained; they are better regarded as the two ends of a continuum.

Metaphor and metonymy often interact in complex ways. One of the most intriguing of these is the fact that a great many metaphors have a metonymic basis: the metaphor MORE IS UP (*A high* number/*Prices* soared) is motivated by the metonymic activation of quantity by verticality, through their frequent experiential association (e.g. in pouring liquids into containers and watching their level rising).

The CTMy has been shown to be highly relevant for the study of cognition and reasoning, lexical semantics, grammar, and discourse. A large number of categories have been shown to have a metonymic *prototype* (a kind of model of the category); see Lakoff (1987). The CTMy provides a useful framework for the study of lexical polysemy, in which metonymy is a fundamental force. But, most importantly, the CTMy has demonstrated the pervasive metonymic motivation of a great many grammatical structures, like raising, conversion, generic sentences, etc., of certain types of pragmatic inferences like indirect speech acts and conversational implicatures, and of discourse strategies.

Further Reading

Barcelona, Antonio (ed.) 2000. Metaphor and metonymy at the crossroads. A cognitive perspective. Berlin and New York: Mouton de Gruyter.

Barcelona, Antonio. 2003. Metonymy in cognitive linguistics. An analysis and a few modest proposals. Motivation in language: studies in honor of Günter Radden, ed. by Hubert Cuyckens, Thomas Berg, René Dirven, and Klaus-Uwe Panther, 223–55. Amsterdam and Philadelphia: John Benjamins.

Gibbs Jr., Raymond W. 1994. The poetics of mind. Figurative thought, language and understanding. Cambridge and New York: Cambridge University Press.

Goossens, Louis, Paul Pauwels, Brygida Rudzka-Ostyn, Anne-Marie Simon-Vanderbergen, and Johan Vanparys. 1995. By word of mouth. Metaphor, metonymy and linguistic action in a cognitive perspective. Amsterdam and Philadelphia: John Benjamins.

Kövecses, Zoltán, and Günter Radden. 1998. Metonymy: developing a cognitive linguistic view. Cognitive Linguistics 9-1. 37–77.

Lakoff, George. 1987. Women, fire and dangerous things. What categories reveal about the mind. Chicago and London: The University of Chicago Press.

Lakoff, George, and Mark Johnson. 1980. Metaphors we live by. Chicago and London: The University of Chicago Press.

Lakoff, George, and Mark Turner. 1989. More than cool reason: a field guide to poetic metaphor. Chicago and London: The University of Chicago Press.

Langacker, Ronald. 1999. Grammar and conceptualization. Berlin and New York: Mouton de Gruyter.

Panther, Klaus-Uwe, and Günter Radden (eds.) 1999. Metonymy in language and thought. Amsterdam and Philadelphia: John Benjamins.

Panther, Klaus-Uwe, and Linda Thornburg. 1998. A cognitive approach to inferencing in conversation. Journal of Pragmatics 30. 755–69.

Ullmann, Stephen. 1962. Semantics. An introduction to the science of meaning. Oxford: Blackwell.

ANTONIO BARCELONA

Mexico

The mosaic of languages in Mexico is among the richest and most complex in the Americas. It is the largest Spanish-speaking country in the world. However, Spanish has only dominated in the region that is today Mexico for the past five centuries. For millennia before, the mountain valleys and tropical forests of Middle America supported a density of population and spectrum of cultural developments that nurtured one of the richest arrays of native languages in the Americas. Mexico, therefore, concentrates the largest number of speakers of one of the most widely spoken languages in the world while preserving the largest number of speakers of the most ancient native American languages.

Demographic and Geographic Profile of Mexican Languages

There are a third of a billion speakers of Spanish in the world so that every one in 18 people in the world speaks this language. In Mexico, there are 90 million first-language speakers of Spanish, almost 90 percent of the country's population of 104 million (estimate, 2002). More than a fourth of all Spanish speakers in the world, therefore, lives in Mexico. Mexico City, the capital, concentrates a population of over 20 million, making it the fourth largest metropolis in the world. It is thereby the largest Spanish-speaking city in the world. Speakers of native languages in Mexico amount to slightly less than 10% of the population. They speak almost 300 different languages. Well over half the population of Mexico is a racial mixture of Spaniard and Amerindian, known as *mestizo*. Just under a third of the population is mainly of Amerindian descent. The speakers of the numerous languages of Mexico occupy an area of more than 750 million square miles. Mexico is four times the size of Spain, where Spanish originated. The area of present-day Mexico, however, is only half of what it was in the nineteenth century. At

that time, it lost much of its territory to what is now the southwestern region of the United States, the area from the states California to Texas. Today, this region has the largest number in the Unite States of first- and second-language speakers of Spanish.

Development and Characteristics of Mexican Spanish

Spanish came rapidly to dominate in Mexico from the earliest colonial times. This dominance resulted not just from the power of the Spanish Conquest from 1519 to 1521. Of most consequence were the devastating diseases brought to the native population by the Spanish. At the beginning of the sixteenth century, just before the Conquest, the native population of Mexico amounted to approximately 25 million. Only three decades later, well over 90% of that population had died from European diseases against which natives had no immunity. The population of Mexico did not again reach its pre-Conquest level until the middle of the twentieth century.

The Conquest had a devastating and permanent effect on the linguistic balance of Mexico. The Spanish-speaking population enjoyed robust growth, its language coming to the forefront in all the colony's commercial, cultural, military, educational, and religious activities. Native languages survived only precariously among geographically and socially isolated remnants of the original native population.

The Spaniards who came to Mexico in the early colonial period amounted to less than 100,000 people. However, they were mostly male and interbred widely with the surviving native women, producing the *mestizo* population. It was this growing population that became the main vehicle for the advancing dominance of Spanish.

Distinctions of Spanish in Mexico

The use of Spanish in Mexico has developed differences from other Hispanic countries in terms of vocabulary, pronunciation, and grammar. These distinctions have occurred based on the historical origins of the language in central and southwestern Spain and its evolution amid the dynamics of Mexican development.

The foundations of Spanish throughout Latin America originated with the language as it was spoken in central and southwestern Spain during the sixteenth and seventeenth centuries. The kingdom of Castile, in the central *mestizo* of the Iberian Peninsula, led the conquest of colonies in the Americas. In addition, soldiers, merchants, administrators, settlers, and missionaries set out from the southwestern areas of Andalusia, primarily the port of Seville, and Estramadura.

The dominance of Castile in the propagation of the language was so extensive that the Spanish language is often also referred to as *castellano* or Castilian. The base of Spanish in Mexico and throughout Latin America reflects patterns dominant in Castile and southwest Spain at the time of colonization. Variations have emerged since then from that base.

While in Spain today the "c" may have a "th" sound, in Mexico it is pronounced now only as an "s." The sound of Castilian "x" at the time of colonization was "sh" but has now evolved in Mexico to a breathed "h" sound. (Galician and Portuguese still maintain the earlier tendency.) The original pronunciation, therefore, of the first syllable of Mexico sounded like "mesh" but has now become "meh". In addition to the way in which speakers of Mexican Spanish tend to abbreviate or eliminate certain sounds and syllables, they also use a distinct musicality and intonation, raising the pitch of a phrase or sentence toward the end. Mexican Spanish is spoken more slowly, musically extending sounds. There is a tendency to emphasize consonant sounds and reduce or "swallow" unaccented vowel sounds.

As the largest Spanish-speaking country, Mexican publishers and media outlets considerably influence Spanish-language communications. This is especially true for Mexican television and radio, which massively penetrates into the Latino population of the United States. Mexican television networks, such as Univision and Telemundo, produce *novelas* (series dramas) that are broadcast around the world. Nonetheless, dominance of Spanish-language media and communications remains with Spain, and, in the Americas, is divided between Mexico and Argentina. The prominence of Mexican literature was recognized in 1990 when Octavio Paz won the Nobel Prize in literature.

The vocabulary of Spanish derives essentially from Latin. However, since the Moors occupied much of the Iberian Peninsula from the Middle Ages until the end of the fifteenth century, words of Arabic origin occur extensively in Spanish. Such is the origin of the often-used Mexican expression, *ojalá* (May it be!). Many words in Mexico beginning with "al" such as *alcalde* (mayor), *algebra* (algebra), and *algodón* (cotton) originate from Arabic.

The most distinctive source for the vocabulary of Mexican Spanish, however, derives from native languages, especially Nahuatl. This was the language of the Aztecs, who ruled the part of Mexico that the Spanish initially conquered and occupied. From the Aztec language, Mexican Spanish has received its words for *chocolate* (chocolate; from the Nahuatl word, *chocolatl*), *tomate* (tomato; from *tomatl*), and *coyote* (coyote; from *coyote*). These words have entered the vocabulary of the entire world. The strongest influence on modern vocabulary development in Mexican Spanish has been English, contributing such words as *béisbol* (baseball), *cóctel* (cocktail), and *sándwich* (sandwich).

As throughout the Spanish-speaking world, the standard, accepted version of the language in Mexico continues to adhere to rules formulated by the Spanish Royal Academy, in Madrid. (Somewhat similar to the way in which French-speaking countries follow norms established by the French Academy, in Paris.) There are, nonetheless, regional Mexican language distinctions. These occur along the border with the United States, around the industrial region of Monterey, the Caribbean coast, southern Amerindian areas, rural regions, and Mexico City. Throughout the latter part of the twentieth century, the United States has increasingly become a center for the speaking of Mexican Spanish due to the economic diaspora of millions of Mexicans to that country.

Beyond Spanish

Less than 10 million Mexicans use the several hundred native languages still surviving in the country. These languages are grouped into 11 linguistic families: Algonquin, Chiapanecan-Mangue, Chinantecan, Huave, Mayan, Mixe-Zoque, Oaxacan, Tarascan, Tlapanecan, Totonacan, and Uto-Aztecan. The major languages still have several hundred thousand to several million speakers.

Colonial missionaries established written forms for the native languages, using the Roman alphabet. Thus, writing today of native languages reflects native pronunciation as transcribed into equivalent colonial Spanish alphabetic sounds. Pre-Conquest written forms were limited to pictograms and/or ideograms. Many of these have only recently been deciphered, and often only partially. Hostility to native languages appeared throughout early Mexican history, their speaking prohibited and many documents in them burned.

The Mayan languages, which in Mexico appear along a southern belt of the country from west of the state of Chiapas then along the border with Guatemala and Belize and then into the peninsula of Yucatan, have nearly two million first- and second-language speakers. The Mayan languages spoken in Mexico are principally Chol-Chontal, Tzeltal, Tzotzil, Mocho, and Yucatecan. In the ancient language, glyphs were used for syllable sounds. The Mayan languages generally follow a syntactical sequence of verb–object–subject. The great literary epic, *Popul-Vuh*, recounting the Mayan version of the creation of the universe did not, however, originate in a Mayan language of Mexico but of Guatemala. Some publishing and radio broadcasting occurs in Mayan languages where there is a more dense concentration of such speakers.

In the southwestern part of Mexico, around the area of Oaxaca, there are half a million speakers of Zapoteco (the "cloud" people) languages. These are highly tonal (i.e. they use different levels of intonational pitch to distinguish words) and require an alphabet of 39 letters to capture their range of sounds. There are under 100,000 speakers of Mixe languages, and a quarter of million speakers of Mexteco languages. This region also has several thousand Chinanteco speakers. Oaxaca and the area around it has the greatest concentration of speakers of native languages in Mexico.

In the central part of Mexico, from Nayarit into the state of Mexico and through Puebla down to Vera Cruz, there are well over a million first- and second-language speakers of Nahuatl (Aztecan) languages. The eastern portion of the region has several hundred thousand speakers of the Totonacan languages. Unlike Mayan languages, Nahuatl does not impose a fixed word order on subject, verb, and object. Moreover, somewhat like German, it is an agglutinative language: Nahuatl strings morphemes together into words that can accumulate into more than ten syllables.

Due to extensive migration from the impoverished areas of southern Mexico into the United States, the native languages of Mexico now have hundreds of thousands of speakers in the latter country. Political and cultural consciousness has given renewed vigor to the native languages. This development is especially true in the state of Chiapas with its armed rebellion against the central government of Mexico. The growth of the *indigenista* movement throughout the twentieth century, beginning with the victory of the Mexican Revolution of 1910, supported extensive anthropologic, archeologic, and linguistic studies of the ancient native languages. Many Mexican and US universities now offer courses in the study of these languages. Publishing and media firms of varying sizes now exist in these languages.

References

Atlas lingüístico de México, Serie estudios de dialectología mexicana dedicada a Pedro Henríquez Ureña, 4. Mexico City: Colégio de México, 1990.

Borah, Woodrow Wilson. 1951. New Spain's century of depression, Ibero-Americana, Vol. 35. Berkeley, CA: University of California Press.

Boyd-Bowman, Peter. 1973. Patterns of Spanish migration to the New World, 1493–1580. Buffalo: Council on International Studies, State University of New York.

Campbell, Lyle, et al. 1978. Bibliography of Mayan languages and linguistics, Albany: State University of New York.

Carlos Castillo. 1987. University of Chicago Spanish-English English-Spanish Dictionary, 4th edition, 1–43, Chicago: University of Chicago Press.

Grace, Lee Ann. 1976. The effect of bilingualism on sixteenth-century Mexican Spanish. Dissertation, State University of New York at Buffalo.

Hill, Jane H., and Kenneth H. Hill. 1986. Speaking Mexicano: dynamics of syncretic language in Central Mexico. Tucson: University of Arizona Press.

Lochhart, James. 1991. Nahuas and Spaniards: postconquest Central Mexican history and philology, Nahuatl Studies Series, 3. Stanford, CA: Stanford University Press.

Marcus, Joyce. 1992. Mesoamerican writing systems: propaganda, myth, and history in four ancient civilizations. Princeton, NJ: Princeton University Press.

Moreno de Alba, José G. 1994. La pronunciación del español en México. Mexico City: Colegio de México.

Perisinotto, Giorgio. 1975. Fonologia del español hablado en la ciudad de México. Mexico City: Colegio de México.

EDWARD A. RIEDINGER

See also **Maya and Mayan Languages; Nahuatl and Uto-Aztecan Languages; Spanish and Iberoromance Languages**

Middle (Classical) Japanese

Middle Japanese is a historical stage in the history of the Japanese language between Old Japanese and Early Modern Japanese. The chronological span of Middle Japanese can be estimated as ninth to twelfth CE, roughly corresponding to the Heian period, although ninth century data in many respects are

transitional from Old Japanese to Middle Japanese, in pretty much the same way as thirteenth century, is transitional between Middle Japanese and Early Modern Japanese. Middle Japanese is a language predominantly based on the dialect of the upper and middle nobility in the capital Heiankyô (present day Kyôto). In contrast to Old Japanese, there is no evidence for any kind of dialectal variation, and the existing sources preserve only bits and pieces of language samples as spoken by commoners and even lesser nobility.

It is necessary to keep in mind that Old Japanese and Middle Japanese do not represent a direct continuity in time, as both of them are based on geographically close but in some respects different dialects. Nevertheless, both can be defined as belonging to the Western dialect group of Central Japanese.

Sources

The written sources of the Heian period composed completely or partially in Middle Japanese can be divided into three major groups: (1) commentaries on Buddhist and Confucian classics; (2) dictionaries; and (3) literary works. The first group of texts, the commentaries, is considerably influenced by Chinese, and mostly uses a cumbersome system of writing that only approximately represents the language. In addition, these texts contain a number of archaisms from Old Japanese, no longer essential for Middle Japanese. The second group, dictionaries, although containing valuable lexical and phonetic materials with clues on vocabulary and pronunciation, respectively, tells us next to nothing about the morphology (word structure and word formation) and syntax (sentence structure) of the language. Therefore, the present paper is based predominantly on the data found in literary texts, although dictionary data are also included for the description of Middle Japanese phonology, and the commentary texts are used for the description of the writing system.

Extant literary texts represent both poetry and prose. The Japanese written language up to the Meiji period (1868–1911 CE) was predominantly based on Middle Japanese, although starting from the Muromachi period there have been texts produced in the vernacular. For this reason, Middle Japanese is often called 'Classical Japanese'.

Writing System

Several writing systems were in use in the Heian period. Literary texts are written in a mixed system including both syllabic signs representing cursive forms of many of (but not all) earlier *man'yôgana* signs and cursive forms of Chinese characters, which roughly stand for stems of words. The syllabic system is polyphonic like *man'yogana*; one syllable can be written by more than one sign. This syllabic system of writing is a direct predecessor of modern *hiragana*, where each syllable is rendered by only one sign. The predecessor of modern *katakana* syllabary also originated in this period, although its usage was mainly limited to Buddhist texts and to indicating the reading of Old Japanese texts written in *man'yôgana*. Only in the twelfth century *katakana* gains somewhat wider usage and penetrates even some literary texts. Finally, there was another writing system with limited usage called *kunten* ('Japanese readings [rendered by] dots'), chiefly used for 'Japanicizing' the reading of texts originally written in Chinese. It was based on placing dots at certain corners or sides of a Chinese character that indicated particles or suffixes of Middle Japanese. As a matter of fact, there were several different *kunten* systems, invented by various Buddhist sects, which usually kept them secret from members of other sects.

Phonology

The sound system of Middle Japanese underwent two major radical changes as compared to the Old Japanese system. The first major change resulted in more than double of the number of consonants when certain consonants (palatalized and labiovelar) were introduced to the system due to the large influx of Chinese loanwords. The second major change was due to the weakening and subsequent loss of certain consonants between vowels that ultimately destroyed the Old Japanese requirement that no two vowels occur adjacent to each other.

Middle Japanese had the following consonants: *-p, f, f', b [^{m}b], b' [^{m}b'], t, t', d [^{n}d], d' [^{n}d'], k, k', k^w, g [^{ŋ}g], g' [^{ŋ}g'], gw [^{ŋ}g^w], m, m', n, n', -N, s, s', z [^{n}z], z [^{n}z'], w, y, r, r'*. Similar to Old Japanese, no consonant is possible in the syllable-final position; thus, only syllables of CV[V] and V structure are found in the native Middle Japanese vocabulary. In Chinese loanwords, however, nasal sonorants *-m, -n, -ŋ* and probably voiceless *-p, -t, -k* as well are possible. Consonants *b, d, g, z,* and r occur only in word-medial position in the native vocabulary. Between vowels, *-f-* shifted to *-w-* and consequently disappeared before the high vowels /i/ and /u/ or went to *-y-* before /e/. The same fate befell the original intervocalic *-w-*. In certain sets of verbs and adjectives, intervocalic *-k-* and *-s-* also disappeared before high vowels /u/ and/or /i/.

The set of Middle Japanese vowels, on the other hand, shrank compared to Old Japanese, as only five

vowels *a, i, u, e, o* are present. The vowels /e/ and /o/ are found in Middle Japanese only in post-consonantal position, that is, they never appear as a word or syllable initial.

From the Chinese–Japanese dictionary *Ruiju Myôgishô* (1081 CE), we have fairly good information on the Middle Japanese accentual system. It was a system that was based on pitch and possessed two major characteristics: register (high (H) or low (L)) and locus (location of the drop in pitch in a given word with a High–Low contour), which are also typical for the modern Kyoto dialect. Like in Modern Kyoto, nominals were characterized by both register and locus, while verbs and adjectives exhibited only register.

Lexicon

There are many loanwords from Chinese in Middle Japanese that penetrated different lexical domains, but they are most numerous among the sociopolitical and religious strata of the Middle Japanese vocabulary: *dairi* 'forbidden city', *kugyau* 'nobility', *sou* 'monk', *kyau* 'sutra'. Many Chinese loanwords have doublets in the native vocabulary, e.g. (native word is given in parentheses): *sekai (yo)* 'world', *kyau (miyako)* 'capital', *guu (miya)* 'palace', *sinwau (miko)* 'prince', *daizin (otodo)* 'minister'. There are also words coined by the Japanese from Chinese elements, e.g. *chûnagon* 'middle councilor', *dainagon* 'senior councilor'. There are also loanwords from Sanskrit borrowed via Chinese or Korean: *butu [but]* 'Buddha' < Middle Chinese *but* < Sanskrit Buddha, *fati* 'bowl for alms' < Middle Chinese *pat* or Old Korean *pati (cf. Middle Korean *pali*) 'id.' < Sanskrit *pa:tra* or *pa:tri* 'bowl'.

Morphology

Middle Japanese morphology underwent some cardinal changes as compared to Old Japanese morphology. First, a number of Old Japanese morphological markers are no longer used in Middle Japanese texts. Second, a number of new morphological markers appear in Middle Japanese that did not exist in Old Japanese, partly due to the fact that Old Japanese and Middle Japanese are based on geographically close but not identical dialects. Third, some constructions of Old Japanese consisting of clearly identifiable word elements underwent the process of reduction in Middle Japanese and conflated into single unanalyzable suffixes , e.g. the Old Japanese phrase *V-am-aku posi* 'wants to do V' became a verbal suffix *-amafosi* (morphological elements and their functions that represent new

developments in Middle Japanese that are not found in Old Japanese are indicated below in bold typeface).

Nominal Morphology

Most nominals (with the exception of some pronouns) in Middle Japanese have no formal markers, distinguishing them from other parts of speech, e.g. *yama* 'mountain', *fito* 'person', *yuki* 'snow', *midu* 'water', *fi* 'fire', *futa* '2', *towo* '10'.

Some Middle Japanese pronouns have two stems: unextended, and extended with suffix *-re*, which may be treated as a formal marker of these pronouns, distinguishing them from other parts of speech, and also from other pronouns that do not have the special stem in *-re*. It is necessary to note that Middle Japanese has a unified system of personal-reflexive pronouns (...self), which does not exist in Old Japanese.

Personal-reflexive pronouns: *wa-/ware* 1/3ps and 1/3pp (rare), **maro** 1ps (attested only once in OJ), **wa-ga mi** 1/3ps, **mi** 1ps, **midukara** 1/3ps, *wono-/wonore* 1/2/3ps and 1/2/3pp, **wono-ga mi** 1ps, **wonodukara** 1/2/3ps, **nandi** 2ps (pejorative), [mi]masi 2ps, **omafe** 2ps (honorific), **gozen** 2ps (honorific), **watakusi** 3ps. Demonstrative pronouns: proximal *ko/kore* 'this', mesial *so-/sore* 'that', distal *ka/kare* 'that over there', **a-/are**. Demonstratives *kore* and *sore* can be used as 3p pronoun 'he' or 'she'. Demonstrative pronouns indicating place or direction: proximal: *koko, konata* 'here', mesial: *soko,* **sonata** 'there', distal *kanata,* **kasiko, anata** 'over there'. Interrogative pronouns: *ta-/tare* 'who', *nani* 'what', *ika,* **ikani, ikaga, ikade** 'how', *iduku/**iduko*** 'where', *iduti* 'where to', *idure* 'which', *itu* 'when', **iku/ikuda**/*ikura* 'how many', *nado/nazo/**nadote*** 'why', **nadeu/nandeu** 'what kind', 'why'. Collective pronouns: *mina,* **subete** 'all'.

In contrast to Old Japanese, only the following native numerals are attested in Middle Japanese: *fito-* 'one', *futa-* 'two', *mi-* 'three', *yo-* 'four', *itu-* 'five', *mu-* 'six', *nana-* 'seven', *ya-* 'eight', *kokono-* 'nine', *towo* 'ten', *fatati* 'twenty', *misoti* 'thirty', *yosoti* 'forty', *musoti* 'sixty', *momo* 'hundred', *ti* 'thousand', and *yorodu* 'ten thousand'. Native numerals above ten have disappeared from the language being replaced by Chinese loanwords, e.g. *zifu iti* [ⁿzip it] 'eleven', literally 'ten' + 'one'. The numerals of Chinese origin below 'eleven' occur only with classifiers of Chinese origin and are never used independently. The system of classifiers is in its infancy, although it is already richer than the Old Japanese system. The following classifiers are well attested: *-tu* (objects used with digits), *-ti* (objects used with tens and hundreds), *-ri* (people), *-fe* (layers and folds), *-ka* (days), **-nin** (persons), **-do** (times), **-ba** (roofs), **-sudi** (long thin

objects), **-tokoro** (high-ranking persons), **-zyau** (wards).

Verbal Morphology

The verbal morphology of Middle Japanese underwent three great changes as compared to Old Japanese. First, in sharp contrast to Old Japanese, all Middle Japanese adjectives behave like verbs and not like nouns. In other words, Middle Japanese adjectives have a comparatively well-developed verb-like paradigm that differentiates between attributive and final forms, e.g. *aka-ki isi* (red-ATTR stone) 'red stone' and *isi aka-si* (stone red-FIN) 'stone is red'. Second, auxiliaries of Old Japanese that followed the infinitive form of verbs became suffixes in Middle Japanese. Third, tense marking has disappeared from the language, while 'retrospective' constructions filled the void. The last two changes can be demonstrated by a single example: old Japanese verb-*i-kyi* verb-INF-PAST became Middle Japanese verb-*iki* verb-RETR.

Middle Japanese verbs are divided into several classes: consonantal verbs, with roots ending in a consonant (*yuk-* 'go', *kir-* 'cut', *omof-* 'love, think', *nokos-* 'leave', etc.), vowel verbs, with roots ending in a vowel (including weak vowel verbs that lose the final vowel under certain conditions like *kofi-* 'love', *tasuke-* 'save', and strong vowel verbs that never lose their root vowel, like *mi-* 'see', *ke-* 'kick', etc.), and irregular verbs (*ko-* 'come', *se-* 'do', *ar-* 'exist', *sin-* 'die', *in-* 'go [away]'). In addition, there are defective verbs: *n-, to* 'be', *to* 'say' that have only a limited number of forms and are irregular as well.

Middle Japanese adjectives are traditionally divided into two classes: one with root ending in *-si* and another ending in any other syllable. The only distinction between these two classes can be seen in the final form, where *-si* adjectives do not take the final marker *-si* to avoid the doubled syllable, e.g. *aka-* 'be red' becomes *aka-si* red-FIN 'is red', but *utukusi-* 'be beautiful' becomes *utukusi* 'is beautiful'. Since very few suffixes can directly follow adjectival stems, adjectives form a special secondary conjugation type, consisting of adjectival infinitive form *-ku* plus the verb *ar-* 'exist' that contracts to *-k-ar-*, e.g.: *aka-ku ar-* be red-INF exist > *aka-k-ar-*, to which negative *-az-* can be further added: *aka-k-ar-az-u* be red-INF-exist-NEG-FIN '[it] is not red'.

Middle Japanese verbs can take both prefixes or preverbs and suffixes, but while a word is limited to only one prefix, multiple suffixes may be added. There are three verbal prefixes, **uti-, kaki-/kai-, and mote-,** with mostly unclear or poorly investigated functions, and one preverb *ye-,* which, in combination with a following negative form of the verb, conveys the lack of ability to perform an action, e.g. *ye-yom-az-u* PREV-recite-NEG-FIN '[he] cannot recite'. There is also one circumfix in Middle Japanese: *na....so,* encircling an infinitive form and making a negative imperative: *na-yak-i-so* NEG-burn-INF-IMP 'do not burn'. Verbal affixes differ in form after consonantal stems and after vowel stems (with the possible differentiation between forms found after weak and strong vowel verbs). Thus e.g. the attributive suffix has the following forms: *-u* after consonantal verbs, *-uru* after weak vowel verbs, and *-ru* after strong vowel verbs: *yuk-u* fito 'person *who goes*', *kof-uru* fito 'person *who loves*', and *mi-ru* fito It is impossible to provide a list of all verbal affixes here, so only the most important are listed: infinitive -*i*/-Ø, negative infinitive -*azu*/-*zu*, finite -*u*, attributive -*u*/-*uru*/-*ru*, imperative -*e*/-*yo*, negative -*an*-/-*az*-/-*n*-/-*z*-, tentative -*am*-/-*m*-, iterative -*af*-, passive- -*are*/-*rare*-, causative/honorific -*ase*-/-*sase*-, debitive -*ubey*-, negative debitive -*umazi*-, retrospectives -*ikyi* and -*iker*-, perfectives -*ite*-/-*it*-/-*te*-/-*t*- and -*in*-/-*n*-, perfective-progressive itar-/-*tar*-, conjectural -*umer*-, optative -*amafosi*/-*mafosi*, desiderative -*abaya*. 'person *who sees*'.

It is impossible to provide a list of all verbal affixes here, so only the most important are listed: infinitive -*i*/-Ø, negative infinitive -*azu*/-*zu*, finite -*u*, attributive -*u*/-*uru*/-*ru*, imperative -*e*/-*yo*, negative -*an*-/-*az*-/-*n*-/-*z*-, tentative -*am*-/-*m*-, iterative -*af*-, passive- **-are/-rare-**, **causative/honorific -ase-/-sase-**, debitive -*ubey*-, negative debitive **-umazi-**, **retrospectives -ikyi** and **-iker**-, perfectives **-ite-/-it-**/-*te*-/-*t*- and **-in-**/-*n*-, perfective-progressive **itar-**/-*tar*-, conjectural **-umer-**, optative **-amafosi/-mafosi**, desiderative **-abaya**.

In addition to this quite complex verbal morphology, there is also the no less complex system of auxiliary verbs that mostly follow the infinitive form of verbs. Auxiliary verbs can be subdivided into several classes (within each class, only the most frequent are listed): honorific (*tamaf-, ofas[e]-, ofos[e]-, imas-*), humble (*tatematur-, tuka[u]matur-, tamafe-, kikoye[sase]-, mawos-*), polite (*faber-, saburaf-*), assertive (*nar-*), cooperative-reciprocal (*af-*), directive (*ide-/ idas-, ire-/ir-, ko-, yuk-/ik-, age-/agar-, kudas-, yar-, watar-, yose-/yor-*), and resultative (*fate-, ok-, tuke-/tuk-, tome-*).

References

Ikeda, Tadashi. 1980. Classical Japanese grammar illustrated with texts. Tokyo: Tôhô Gakkai.
Takeuchi, Lone. 1987. A study of classical Japanese tense and aspect. Copenhagen: Akademisk Forlag.
Vovin, Alexander. 2003. A reference grammar of classical Japanese prose. London and New York: Routledge/Curzon.

ALEXANDER VOVIN

See also: **Japanese; Old Japanese**

Middle English

Middle English is the name given to the varieties of English spoken between the Norman Conquest of 1066 and the introduction of the printing press to England by William Caxton in 1476. These historical events are simply convenient demarcation points; in reality, many of the linguistic changes characterizing Middle English started before and extended beyond these dates.

Background

In 1066, William, Duke of Normandy, also known as William the Conqueror, invaded England, defeated the English armies at Hastings, and declared himself King of England. Throughout the Middle English period, England was ruled by royals who were of French descent: the Norman line from 1066 to 1154 and the Plantagenet line from 1154 to 1485. The demographic effect of the Norman Conquest was not overwhelming, and people who spoke only French probably never exceeded 2% of the total population. The social and cultural consequences of this event, however, were substantial. The positions of power and prestige and the centers of learning and writing passed from Anglo-Saxon to Norman hands. Virtually every aspect of the history of the language during Middle English has to be described and studied both on its own terms and in the context of continuing bilingualism and imitation of Continental sources.

Early Middle English usually covers the time from the mid eleventh to the mid thirteenth centuries. Relatively few records written in English survive from that period; most of the historical documentation and literary production during the Early Middle English period was in Latin or in Anglo-Norman, the variety of French used by William the Conqueror, his successors, and their courts. For English, it was a time of rapid reduction of grammatical markers and borrowing of vocabulary from Anglo-Norman. Central Middle English stretches roughly from the mid thirteenth to the mid fourteenth centuries, up to the beginning of the Hundred Years' War (1337–1453) and the epidemic of bubonic plague known as the Black Death (1348–1350). The Central Middle English period is characterized by further attrition of inflections, vowel length adjustments, and the continuing introduction of new loanwords from French, although the use of French was already on the decline. Late Middle English is the time during and

after Chaucer's life (born c. 1340–1346, died 1400), up to the introduction of printing. (Caxton brought out the first printed edition of *The Canterbury tales* in 1478.) It was during the Late Middle English period that the pronunciation changed in a complex process commonly referred to as the *Great Vowel Shift*. Furthermore, when vowels occurred at the end of a word and were unstressed, they were completely lost in all dialects, and since word endings often had grammatical significance, the grammar changed as well. Late Middle English also marks the beginning of standardization, the complete replacement of French by English as the official, legal, and parliamentary language, and the rise in literacy.

Dialects

One consequence of the political and cultural dominance of the Anglo-Normans was that during the first three centuries of Middle English, no regional variety of the language achieved prestige and prominence comparable to that of West Saxon during the Old English period. Each geographical area had its own scribal and literary traditions. Dialects differed both in sound structure and vocabulary, as well as in word and sentence structure. The main north–south divisions of Middle English are known as Northern, Midland, and Southern Middle English. Within the south, a separate dialect, Kentish, developed in the counties of Kent and Sussex. The Midland dialect area splits further into two linguistically distinct varieties: East Midland and West Midland.

Orthography

Another consequence of the Norman Conquest was the replacement of the 'Insular' script used by the Anglo-Saxons with a Continental form of the Roman alphabet known as 'Carolingian' script. The letters <æ> and <ð> were preserved only in the earliest texts. The Old English yogh <ȝ> was used to represent both [j], as in *ȝow* 'you,' and [ç, x], as in *heiȝ (e)* 'high,' *rouȝ(h)* 'rough'; the continental <g> was adopted for [g] and [ǰ]. The letter thorn <þ> was very similar in shape to the letter <y>; hence, y^e, y^t stand for *the, that*. (y^e, y^t were retained in printers' fonts during the fifteenth and sixteenth centuries, which gave rise to pseudo-antiquarian spellings like *Ye Olde Antique Shoppe*.) New digraphs (two-letter combinations)

were introduced for the spelling of many long vowels and diphthongs: <ou, ow> for [u:], <ee> for [e:], <oo> for [o:], etc. To avoid the visual confusion of adjacent vertical strokes, the letter <u> was replaced by <o> when it was next to <m, n, u, v, w>. Thus, *sume* 'some', pronounced [sʊm(ə)], *huni* 'honey', pronounced [hʊni], and *luue* 'love', pronounced [lʊv(ə)], were respelled to *some, hon(e)y, loue*. New consonant letter combinations introduced by the Anglo-Norman scribes were: <ch, cch, tch> for [č], <gg(e)> for [ǰ], <gh> for [ç, ɣ, x], <wh, qu, quh> for earlier <hw>, and <sh, ssh, ss> for [ʃ].

Sound System

Middle English was a time of rapid change of the quality and the quantity of many vowels. The southern dialects of Late Middle English, which served as the basis of the emerging fifteenth-century standard, had five short vowels and seven long vowels:

Short vowels		Long vowels	
I	ʊ	i:	u:
		e:	o:
ɛ	ɔ	ɛ:	ɔ
	a		ɑ:

The colon marks vowel length. In dictionaries and edited manuscripts, length may be indicated by a macron (¯) over the vowel letter. Whether [ɑ:] or [ɔ:] is used in a particular word is an important dialect criterion: in the south, Old English [ɑ:] was raised and rounded to [ɔ:], while the northern dialects retained the low vowel: compare Northern Middle English *stan(e), mair* with Southern Middle English *ston(e), more*, which developed from Old English *stān* 'stone', *māra* 'more'.

Middle English inherited or developed a range of diphthongs that were to become an important component of the changes included in the *Great Vowel Shift*. The diphthongs of Late Middle English were: [ij] as in *stil(e)* 'stile', [ej] as in *weʒ* 'way', [ew] as in *newe* 'new', [iw] as in *stiward* 'steward', [uw] as in *ful* 'fowl', [ow] as in *bowe* 'bow', and [aw] as in *lawe* 'law'. Toward the end of the period, the long vowels /i:, e:, u:, o:/ had already started changing to /əj, ij, əw, uw/ in the south.

Middle English continued the Old English tendency to make vowel length dependent on its environment. A change started in the north toward the end of the twelfth century, and gradually reached the south by the end of the fourteenth, whereby stressed vowels in open syllables (syllables with no final consonant) were lengthened in words of two syllables. Examples of this development are:

Old English *talu*	[talə] 'tale'	Middle English *tale*	[ta:l(ə)]
Old English *nosu*	[nɔzə] ' nose'	Middle English *nose*	[nɔ:z(ə)]
Old English *brecan*	[brɛkən] 'break'	Middle English *brek(e)*	[brɛ:k(ə)]

The lengthening often did not take place when the second, unstressed, syllable of the word was not lost. This produces patterns of historically related words like *game ~ gammon, late ~ latter, shade ~ shadow, throat ~ throttle, nose ~ nostril*. An important spelling consequence of this change is that, in the modern language, words that contain the orthographic sequence -VC*e* (any vowel + any consonant + final -*e*) are usually interpreted as containing a long (or tense) vowel.

The opposite tendency, toward shortening, was also a continuation of Old English changes. Long vowels were shortened before two or more consonants, thus: Old English *gōdspell* 'gospel'> Middle English [gɔsp(ə)l], Old English *crēpte* [kre:pte] 'crept' > Middle English *crepte* [krɛpt(ə)]. These shortenings are responsible for Modern English patterns like *cheap ~ Chapman, dear ~ darling, five ~ fifth, wise ~ wisdom*. This change is also connected with the modern spelling convention, whereby consonants are doubled when adding suffixes only after short vowels, as in *beg ~ begged, drum ~ drummer, fat ~ fatter, sit ~ sitting, red ~ redden* (compare *fate ~ fated, site ~ siting*).

The most far-reaching single event in Middle English sound change, which profoundly affected the general shape of the language, was the loss of final unstressed vowels. In early Middle English, final unstressed vowels lost their qualitative distinctions and merged into the mid-central vowel schwa [-ə], usually spelled <-e>. By c. 1400, all final schwas, including a number of grammatical endings, had disappeared from the pronunciation; the <-e> remained in the language as a spelling convention indicating the quantity of the vowel in the preceding syllable, e.g. *bite ~ bit, cut ~ cute, mate ~ mat, mete ~ met, note ~ not*.

During Middle English, the consonantal system reached its present-day state except for the later (seventeeth century) introduction of /ʒ/ as in *measure*. Long consonants, as in *bedd* 'bed' and *sterre* 'star,' were simplified. A combination of factors, most notably the influx of French loanwords, resulted in an enriched system of consonant distinctions. The voiced fricatives, [v, ð, and z], which in Old English could appear only in the middle of a word, when flanked by vowels, became independent phonemes, allowing minimal pairs such as *fast - vast, thigh - thy*, and *seal* 'acquatic mammal' - *zeal*. The consonant /h/ remained

stable only initially before vowels. The chart below shows the consonant inventory of late Middle English:

	Labial	Dental	Alveolar	Palatal	Velar
Voiceless stops	p	t			k
Voiced stops	b	d			g
Voiceless fricatives	f	θ	s	ʃ	h, x
Voiced fricatives	v	ð	z		
Affricates			č, ǰ		
Nasals	m	n			
Liquids		l, r			
Approximants				j	w

Word and Sentence Structure

The weakening and loss of final unstressed vowels in Middle English resulted in a very significant reduction of distinctive grammatical markers. Until the twelfth century, English nouns and their determiners agreed in grammatical gender: *se dæȝ* 'the day' (masc.), *seo spræc* 'the speech' (fem.), and *þæt word* 'the word' (neuter). The Old English gender system was grammatical, because the fact that a noun was feminine or masculine did not necessarily reflect feminine or masculine traits of the object referred to. Most of the gender marking was lost during the Middle English period, and for the few elements that still show gender, such as the pronouns *he, she, it,* the gender reflects actual sex differences. While Old English inflected nouns differently depending on whether they occurred as subjects or objects, this distinction was also gradually leveled out and had disappeared completely by the end of the Middle English period. The plural and the possessive *–es* endings, used only for one subgroup of masculine and neuter nouns in Old English, spread to all nouns. The difference between singular and plural adjectives (marked by *-e* for the plurals) was abandoned after the end of the fourteenth century.

The personal pronouns in Middle English had the same distinctions as in Modern English: singular/plural, subject/object, first/second/third person, and three genders for the third-person singular. The most important innovation in the pronoun system of Middle English was the adoption of the Scandinavian pronouns *they, them, their,* which replaced Old English *hie, hem, hire.* Throughout Middle English, there were two separate second-person pronouns: *thou, thee, thine/thy* (singular) and *ye, you, your* (plural).

Verb forms in Early Middle English were marked for number, person, tense, and mood. Many of the inflections were identical, however: *-e* was the first-person singular present tense, the singular subjunctive for all persons, present and past tense, the singular imperative, etc. The phonetic loss of unstressed vowels resulted in essentially the modern system except for the survival of *-(e)st* and *-(e)th* for the second- and third-person singular, present tense. The modern third-person singular present tense suffix *-(e)s*, an inflection of obscure origin, possibly influenced by Scandinavian, was originally restricted to northern texts. It spread to the Midlands and the south during the fourteenth century and reached London by c. 1400. Many formerly 'irregular' verbs began to develop regular past tense forms ending in *-ed: cleave ~ clove ~ cleaved, help ~ holp ~ helped, leap ~ lope ~ leaped, walk ~ welk ~ walked.*

During the Middle English period, many complex verb forms and auxiliary verbs continued to develop and enrich the verbal system. Except for the passive progressive forms (*is/was being written*), all other progressive forms, the perfect, and the pluperfect forms (*has arrived, had arrived*) were in use by Late Middle English. The basic modal auxiliary verbs, *can, may, must*, were also in place, and the earlier modal auxiliaries *shall* and *will* were often used as pure future tense markers. The verb *do* was used as a tense marker, a causative verb, and a substitute verb, but not as an auxiliary verb in its modern functions. Questions were formed by simple inversion of the verb and the subject: *Seist thou hit me?* 'Are you saying it to me?' Negation in the early texts could involve attaching the particle *ne* 'not' before the verb, but gradually the negative function was taken over by the originally emphatic *naht* 'naught, not' from *nawiht* 'nothing', often positioned immediately after the verb. Multiple negation was common: *Of fule spaches hie ne mai nauht þolien.* '(Of) foul speaking they not can/may not tolerate.'

In the course of Middle English, the order of the sentence elements became essentially identical to that of Modern English, Subject–Verb–Object, in both main and subordinate clauses. Fixing the position of the subject and the object with respect to the verb is related to the loss of grammatical markers. Unlike Modern English, where only certain adverbs such as *ever, never,* and *sometimes* can intervene between an auxiliary and its main verb, Middle English allowed the object to be enclosed between the two parts of the verb: *I ne* **haue** *nowt but mi swerd* **ibrout**. 'I not-**have** nothing except my sword **brought**.'

Vocabulary and Word Formation

The most serious linguistic consequences of the Norman Conquest were in the area of vocabulary and word formation. It is estimated that approximately 10,000 words were borrowed from Anglo-Norman and

French into Middle English. Seventy-five percent of these early loanwords are still in use. Among them are (a) everyday words: *air, beast, close, dangerous, feast, flower, jealous, journey, liquor, mountain, noble, river, soil, tender, very,* etc.; (b) legal, administrative, and political terms: *army, assembly, council, defense, judge, liberty, navy, parliament, record, servant, soldier, tax;* and (c) words from the spheres of literature, art, science, and medicine: *beauty, color, grammar music, poet, physician, romance, surgeon.* New words continued to be produced by compounding: *bedtime* (c. 1250), *hangman* (c. 1393), *coal-black* (c. 1250), *to overcast* (c. 1225), and *worldly-wise* (c. 1415) are some examples, with the earliest recorded dates in parentheses. Mixed-source compounds began to appear too: *breast-plate* (1386), *freemason* (1376), and *knight erran*t (1350), combining English with French roots, while in *commonweal* (1330), *cornerstone* (1300), and *gentleman* (1275), the French root is followed by an English root. The inventory of Old English prefixes and suffixes was increased by a large number of Romance ones: *en-, mis-, re-, sub-, able-, -ance, -ess, -ment, -ous*, which could produce mixed-origin words such as *talkative, unknowable, wizard* (English roots + Romance suffixes), and *colorless, cheerful, spousehood* (Romance roots + English suffixes). Related to the fact that Middle English lost final unstressed vowels was the new tendency of freely converting words from one part of speech to another, i.e. nouns could simply be used as verbs, verbs as nouns, adjectives as verbs, etc., without any particular marking. Examples of this process are the verbs *child* 'give birth' (c. 1200), *calm* (c. 1399), *cripple* (c. 1300), *tame* (c. 1315), and *word* 'speak' (c. 1205), all formed from earlier nouns and adjectives.

References

The richest ME database is the Middle English Dictionary, based on the analysis of a collection of over three million citation slips. Web access is available through the electronic Middle English Compendium, which includes also a hyper-bibliography of Middle English prose and verse and an associated network of electronic resources.

Blake, Norman (ed.) 1992. The Cambridge history of the English language: Vol. II, 1066–1476. Cambridge: Cambridge University Press.

Brunner, Karl. 1887/1963. An outline of Middle English grammar, translated by Grahame Johnston. Oxford: Blackwell.

Fisiak, Jacek (ed.) Studies in Middle English linguistics. Berlin and New York: Mouton de Gruyter.

Jones, Charles. 1972. An introduction to Middle English. New York: Holt, Rinehart & Winston.

Jordan, Richard. 1925/1970. Handbook of Middle English grammar: phonology, translated by Eugene Joseph Crook, The Hague: Mouton.

Kurath, Hans, Sherman M. Kuhn and Robert E. Lewis (eds.) 1952–2001. Middle English dictionary. Ann Arbor: University of Michigan Press and London: Oxford University Press. Electronic access: http://ets.umdl.umich.edu/m/med/.

McIntosh, Angus, M.L. Samuels, Michael Benskin, with the assistance of Margaret Laing, and Keith Williamson. 1986. A linguistic atlas of late mediaeval English, 4 vol. Aberdeen and New York: Aberdeen University Press.

McSparran, Frances (ed.), 2002. Middle English Compendium. Ann Arbor: University of Michigan press. Electronic access: http://ets.umdl.umich.edu/m/med/.

Minkova, Donka. 1991. The history of final vowels in English, Berlin and New York: Mouton de Gruyter.

Mossé, Fernand. 1968. A handbook of Middle English, translated by James A. Walker. Baltimore: Johns Hopkins Press.

Mustanoja, Tauno F. 1960. A Middle English syntax. Helsinki: Société Néophilologique.

Wright, Joseph, and Elizabeth Mary Wright. 1928. An elementary Middle English grammar, 2nd edition. London: Oxford University Press, Reprint 1967.

DONKA MINKOVA

Migration and Language

In some sense, probably all the world's languages are migrants. In a few cases, languages are spoken in a particular region because the ancestors of their speakers were the first people ever to live there: thus, Austronesian languages were spread across previously uninhabited Oceanic islands between about 1600 BCE and 1300 CE. Often, however, languages are carried by new settlers into a region already inhabited by speakers of different languages: hence, for instance, indigenous languages of the Americas and Australia have been largely replaced by Western European languages during the last few centuries. Such replacements may occur gradually if the migrants have techniques of food production that enable them to outreproduce the original inhabitants (*demic diffusion* or the *wave of advance* model): thus, Bantu-speaking farmers spread over much of southern Africa between about 3000 BCE and 500 CE, into territory previously

occupied by hunter–gatherers and herders speaking non-Bantu languages. Alternatively, a relatively small but highly organized group may impose a new language on an existing population through military conquest (*élite dominance*), as seems to have occurred in the Roman empire, giving rise to the Romance languages.

Migration may have a range of linguistic consequences of great theoretical interest. If it results in two languages coming into contact, one may die out because of deliberate suppression or because the speakers choose to shift to the other language on socioeconomic grounds. If both remain in existence, there may be lexical borrowing, particularly into the language perceived as having less prestige (thus, the Roman occupation of Britain resulted in extensive borrowing of vocabulary from Latin into Brittonic Celtic). In cases of long-term stable multilingualism in a community, the languages may converge in structure, as has been observed in Kupwar in India between Indo-European and Dravidian languages. Finally, a pidgin may arise when speakers of different languages have some strong need to communicate, but cannot for some reason acquire each other's languages fully, as, for instance, in trading situations or on slave plantations in the colonial era.

Another particularly significant result of migration is the creation of language families. When a speech community splits because of movement of some speakers to new areas, the language spoken by the different groups will diverge as various changes accumulate in each population (resulting from internal drift or external influence) and are unable to propagate to the other populations. Over time, this diversification produces a "family tree" of languages (the Austronesian languages seem to provide a particularly clear-cut example). However, if contact is maintained between the speech communities, changes will be able to diffuse between neighboring groups, producing a dialect continuum of related speech forms, rather than a clearly defined tree (as is still seen to some extent in Western Europe within parts of the Romance and Germanic groups).

Such language families provide the basis for comparative historical linguistics. However, their existence has also been used by several nonlinguists as evidence for prehistoric population movements. Thus, from an archeological perspective, Colin Renfrew has suggested connecting the Indo-European family with the spread of farming from Anatolia across Europe and India after approximately 6500 BCE and linking the controversial Nostratic language grouping with still earlier movements from the Fertile Crescent. Luigi Cavalli-Sforza has noted that genetic discontinuities often correspond to linguistic divisions: for instance, Rh-negative individuals are considerably more common among the Basques than other Europeans, and Basque is a linguistic isolate, unrelated to its Indo-European neighbors. He argues that the Basques may represent a pre-Indo-European people, confined to a small mountainous location by later population incursions.

Besides the mere existence of language families, other linguistic information can contribute to an understanding of early migrations. According to the *age–area hypothesis* (formalized by Isidore Dyen as *migration theory*), the location of greatest diversity within a language or family is likely to be the original homeland from which the family spread, because the deepest divisions reflect the earliest splits; moreover, this theory minimizes the number of movements required to explain the current distribution. Thus, English dialectal diversity is greater within Britain than in the more recently settled English-speaking parts of the world. Linguistic paleontology (the study of reconstructed cultural vocabulary) may also provide clues to the area of origin of a family. A third line of evidence is provided by borrowed vocabulary, revealing contact with other groups: the movements of the Gypsies from India to Iran, the Caucasus, Anatolia, and then to the Balkans have been traced by Kaufman in various sets of loanwords in Romani.

A recent strand of research relating to migration and language has considered the distribution of linguistic diversity around the world. Johanna Nichols, drawing on techniques of the population sciences, has sought to quantify structural and genetic diversity of languages in different areas and has distinguished between *residual zones* (characterized by high genetic density and structural diversity, often in mountainous regions serving as linguistic refuges: the Caucasus is an example) and *spread zones* (whose low genetic and structural variation results from the spread of a single language or family across the area: an instance is the Eurasian steppe). She has also argued that the high diversity in the Americas implies earlier and more numerous settlements than those proposed by Joseph Greenberg, whose 'Amerind' people are claimed to have entered the Americas approximately 14,000 years ago. Daniel Nettle, however, adopting Robert Dixon's evolutionary–biological model of *punctuated equilibria* for language change, has opposed her theory that languages ramify at a constant rate over time. He proposes that the initial movement of people into a large uninhabited area (such as the Americas) would promote rapid diversification as communities occupied available habitat niches, followed by gradual areal convergence or local language replacement, resulting in a decline in

linguistic and genetic diversity. A Neolithic transition in an area would provoke a faster reduction of diversity, as farming groups advanced through previous hunter-gatherer territory (as in Eurasia and Africa, where initial human settlement and development of food production were comparatively early).

References

Campbell, Lyle. 1998. Historical linguistics: an introduction. Edinburgh: Edinburgh University Press.

Cavalli-Sforza, Luigi. 2000. Genes, people and languages. London: Allen Lane.

Dixon, Robert. 1997. The rise and fall of languages. Cambridge: Cambridge University Press.

Dyen, Isidore. 1956. Language distribution and migration theory. Language 32.

Greenberg, Joseph. 1987. Language in the Americas. Stanford: Stanford University Press.

Nettle, Daniel. 1999. Linguistic diversity. Oxford: Oxford University Press.

Nichols, Johanna. 1990. Linguistic diversity and the first settlement of the New World. Language 66.

Nichols, Johanna. 1992. Linguistic diversity in space and time. Chicago and London: University of Chicago Press.

Renfrew, Colin. 1987. Archaeology and language: the puzzle of Indo-European origins. London: Jonathan Cape.

Renfrew, Colin, 1999. Nostratic as a linguistic macrofamily. Nostratic: examining a linguistic macrofamily, ed.by Colin Renfrew and Daniel Nettle. Cambridge: McDonald Institute for Archaeological Research.

Sapir, Edward. 1949. Time perspective in Aboriginal American culture: a study in method. Selected writings of Edward Sapir in language, culture and personality, ed. by David Mandelbaum. Berkeley: University of California Press.

MARISA LOHR

See also **Archeology and Language; Genetic Relationship; Pidgins and Creoles**

Miskito and Misumalpan Languages

Miskito is the most widely spoken language of a Central American language family known as Misumalpan. John Mason coined the term Misumalpan using the first two letters of its three branches: Miskito, Sumo, and Matagalpan. Miskito and Sumo are spoken in the Mosquito Coast region of Eastern Nicaragua and Honduras by approximately 150,000 and 8,000 people, respectively. The now extinct Matagalpan languages, Matagalpa and Cacaopera, were spoken in the Central Highlands of Nicaragua and Southeastern El Salvador.

Misumalpan is a relatively small language family that in the pre-Columbian period was surrounded geographically by larger and more widespread language families from Mesoamerica (Otomanguean, Uto-Aztecan, and Mayan) and lower Central America (Chibchan). Before the ninth century, Misumalpan languages are presumed to have been spoken in all of the three major ecological zones of Nicaragua: (1) the Pacific Lowlands, (2) the Central Highlands, and (3) the Caribbean Coastal Plain. However, starting at this time, two major waves of migration from Mesoamerica resulted in Mesoamerican languages displacing Misumalpan languages to the East.

Speakers of the Otomanguean language family who originated in Soconusco, Chiapas comprised the first wave that settled in the Pacific Lowlands and the

Nicoya Peninsula of modern Costa Rica. These migrants were known by the Spanish as the Chorotega and they left behind two now extinct languages—Mangue and Subtiaba. Uto-Aztecan migration occurred slightly after the Otomanguean migration. The first Nahau-speaking people arrived in the ninth and tenth centuries and settled in what is now El Salvador and Northwestern Nicaragua. The second wave of migrants, whose languages came to be known as Pipil or Nicarao, arrived in the twelfth century and settled on the Western shore of Lake Nicaragua.

By the time of the Spanish conquest, Misumalpan languages had virtually been eliminated from the Pacific Lowlands. However, the western branch of Misumalpan languages continued to be widely spoken in the Central Highlands. Matagalpa, also known as Pantasmas, survived throughout the colonial period but by the end of the nineteenth century it had been replaced by Spanish. Nevertheless, an Indian ethnic identity and certain communal institutions and land tenure systems continue into the present in the Nicaraguan province of Matagalpa. In the Morazán province of Southeastern El Salvador, Cacaopera survived into the twentieth century but at present, it has been completely replaced by Spanish. The case of Cacaopera has puzzled linguists because it occurred well out of the Nicaraguan homeland of Misumalpan

languages. Together, Matagalpa and Cacaopera comprise Matagalpan, the name used to refer to the western branch of Misumalpan languages.

Although several distinct dialects of Miskito exist, Miskito is more homogenous than Sumo and Matagalpan and it is therefore considered an isolate within the language family. Sumo and Matagalpan are more internally differentiated and are also more closely related to one another than either is to Miskito. Therefore, they are grouped by linguists in a subfamily called Sumalpan. In turn, Sumo has two distinct branches—Northern Sumo and Southern Sumo. Three extant dialects of Northern Sumo, Tuahka, Panamahka, and Tawahka, continue to be spoken today while Southern Sumo has a single extant representative, Ulwa, and other extinct dialects such as Kukra. Tuahka, also spelled Twahka, is spoken by a few hundred people in and around the village of Wasakin along the Bambana River in the Mosquito Coast. Panamahka is the most common Sumo dialect and it is spoken by about 6,000 speakers living in the area of the Waspuk River and its tributaries in Nicaragua. Tawahka is spoken in Honduras by about 1,000 people around the Patuka River. Finally, the Southern Sumo language Ulwa is spoken in the village of Karawala on the Rio Grande de Matagalpa by about 350 people. Ulwa is in the final stages of extinction as residents of Karawala, most of whom define themselves as ethnically Ulwa, are becoming Miskito, English, and Spanish speakers, as opposed to the long-standing pattern of Miskito-Ulwa bilingualism.

Whereas Sumo and Matagalpan languages have geographically receded over the last 1,000 years, Miskito expanded during the colonial period. Believed to have originated in the Cape Gracias, a Dios region at the terminus of the Coco River that forms the eastern border between Nicaragua and Honduras, Miskito is now spoken by over 100,000 people living along the coast and major rivers between the Escondido River in Nicaragua and the Black River in Honduras. As a result of the dislocations caused by the Contra-Sandinista War, Miskito speakers now live in large numbers in Managua, Miami, and Port Arthur, Texas. In alliance with the British who became commercially and militarily active in the region starting in the 1640s, Miskito-speaking people became a strong force in the Western Caribbean allowing them to expand their area of influence and settlement well beyond their traditional boundaries. As a result, Miskito place names can be found as far as the Caribbean coastal areas of Costa Rica and Panama. Hence, the Sixaola River that divides Costa Rica comes from the Miskito 'Siksa' (Black) and 'Awala' (River).

This Miskito historical ascendance explains in part the relative homogeneity of Miskito as well as its extensive borrowing of vocabulary from English and West African languages, whose influence was brought to the region as a result of English slavery and subsequent West Indian migration to coastal Central America precipitated by banana, mining, and lumber businesses. While Miskito also borrows from Spanish, linguists have found that Sumo and Matagalpan languages demonstrated a clear preference for borrowing from Spanish, while Miskito has historically borrowed primarily from English. This undoubtedly results from the coastal and external trade orientation of the Miskito as opposed to the upriver and subsistence agricultural pattern of the Sumo, who live on the eastward-moving frontier between Spanish-speaking Western Nicaragua and the Mosquito Coast. In a volume about Mosquito Coast Creole English, Johm Holm extensively studied the influence of Mosquito Coast English and West African languages on Miskito. He concluded that Miskito grammar and phonology was not significantly influenced by African languages but lexical borrowing from among others, Fon, Twi, Bambara, Igala, Yoruba, and Kongo, is unmistakable. For example, Miskito people tell 'anansi' stories (from the Twi word for spider) and their sorcerers are known as 'sukias' (from the Fon word sukuna). Whereas the Sumo word for river mouth is 'tikipas', the Miskito word is 'bila'—derived from the Kongo 'bwiila'. Mosquito Coast English and Miskito are deeply interpenetrated. Unlike Sumo and Matagalpan languages, Miskito was able to project words into Caribbean Creole Englishes spoken in Belize, Jamaica, and elsewhere as well as British English itself. For example, the English word dory (canoe), which has an Oxford English Dictionary entry, entered British English through Miskito.

The ascendancy of Miskito people also explains in part the structural similarities between Misumalpan languages. In other words, linguists have noted the anomalous fact that based on the large degree of lexical differentiation, Misumalpan languages have clearly been separated from a common ancestor for a long time. In fact, based on the method of glottochronology, Morris Swadesh estimated that 43 minimum centuries have passed since they diverged. Yet, despite their distant point of divergence, they possess remarkably similar grammatical structures, which at first glance would suggest a more recent common origin. With regard to the similarities in grammatical structure between Miskito and Sumo, the widespread and historically deep pattern of multilingualism in the Mosquito Coast explains in part this incongruity. The majority of Sumo speakers, who have historically been in a subordinate position to

Miskito speakers, also speak Miskito as a second language.

The specific nature of multilingualism in the Mosquito Coast is directly related to the noted ethnic hierarchy and division of labor in the region that developed in the late nineteenth and early twentieth centuries when North American banana, lumber, and mining industries operated heavily in the area. During this time, English-speaking Creoles, Afro-Nicaraguans who emphasize their cultural and social connections to the West Indies, were the dominant locals who occupied the highest status occupations in the enclave economy of the region. Miskito and Spanish-speaking laborers from the Pacific occupied the next rung followed by Sumo Indians at the bottom of the hierarchy. Susan Norwood has noted a sociolinguistic landscape in which the lower status groups tend to learn higher status languages in addition to their own langauges. Hence, Miskito speakers typically learn Creole English and/or Spanish, while Sumo speakers tend to learn Miskito and Creole and/or Spanish when possible. Miskito speakers rarely learn Sumo, and English speakers often have some proficiency in Miskito but none in Sumo. Spanish speakers have by far the highest degree of monolingualism. Norwood found that some Northern Sumo speakers in the Ulwa village of Karawala spoke Panamahka, Ulwa, Miskito, English, and Spanish. Such situations of intense multilingualism and social stratification have had clear linguistic consequences, including the aforementioned structural similarities.

In 1920, German linguist Walter Lehmann convincingly demonstrated the unity of Misumalpan languages. Since then, linguists have speculated about its relationship with other language families. Until recently, linguists have commonly accepted the so-called 'Macro-Chibchan hypothesis', that postulates that Misumalpan and Chibchan languages share a common origin. This contention is based on an evaluation of the distribution of Chibchan languages in lower Central America as well as comparisons of extant Chibchan languages to Misumalpan languages. Recently, linguists such as Ken Hale, Danilo Salamanca, and Lyle Campbell have withdrawn their support from the inclusion of Mismulpan languages in Macro-Chibchan based on the inconclusiveness of the evidence.

Arguments for a Chibchan connection point to the presence of two clearly Chibchan languages both to the North and South of the Misumalpan heartland. Paya, also known as Pech, is a Chibchan language spoken by about 300 people in southern Honduras. Rama, spoken by about 25 people on Rama Cay close to Bluefields, Nicaragua, has been identified as part of the Votic branch of Chibchan that includes Costa Rica's Guatuso. Collete Craig has intensively studied and sought to preserve this language, which has now been almost completely replaced by Creole English. Adolfo Constenla Umaña, a Costa Rican scholar, has made a well-documented case for the plausibility of a shared common ancestor between Chibchan and Misumalpan. In an ambitious synthetic effort to categorize languages of the Americas, Joseph Greenberg presents 226 etymologies in his case for a high-level 'Chibchan-Paezan' family that includes Misumalpan as well as distant languages that have not been previously thought to be related such as Yanomami, Tarascan, and Chimu.

With the exception of Miskito, Misumalpan languages are in danger of becoming extinct in the near future. Despite the tireless efforts of Nicaraguan and North American linguists such as Collete Craig for Rama and Ken Hale and Tom Green for Ulwa, both these languages are not likely to last another generation. Sumo, whose Northern variant, is now sometimes called Mayangna, stands in a better position, particularly in light of recent legal victories for Sumo communities in Nicaragua.

References

Campbell, Lyle. 1997. American Indian languages: the historical linguistics of native America. New York: Oxford University Press.

Constenla Umaña, Adolfo. 1990. Una hipótesis sbore la localización del protochibcha y la dispersión de sus descendientes. Filología y Lingüística (San Jose, Costa Rica) 13. 129–61.

Craig, Colette, and Kenneth Hale. 1992. A possible Macro-Chibchan etymon. Anthropological Linguistics 34. 173–201.

Green, Thomas. 1999. A lexicographic study of Ulwa. Ph.D. Dissertation. MIT.

Greenberg, Joseph. 1987. Language in the Americas. Stanford: Stanford University Press.

Hale, Ken, and Danilo Salamanca. 2002. Theoretical and universal implications of certain verbal entries in dictionaries of the Misumalpan langauges. Making dictionaries: preserving indigenous languages of the Americas, ed. by William Frawley, Kennetch Hill and Pamela Munro. Berkeley: University of California Press.

Holm, John. 1978. The Creole English of Nicaragua's Miskito Coast: its sociolinguistic history and a comparative study of its lexicon and syntax. Ph.D. Dissertation. University of London.

Lehmann, Walter. 1920. Zentral-Amerika, Teil I, Die Sprachen Zentral-Amerikas in ihren Beziehungen zueinander sowie zu Sud-Amerika und Mexiko. Berlin: Dietrich Reimer.

Mason, John. 1940. The native languages of Middle America. The Maya and their neighbors, ed. by Clarence Hay. New York: Appleton Century Company.

Norwood, Susan. 1993. El Sumo, Lengua Oprimida. Wani 14. 53–64.

BARON PINEDA

Modern Linguistics

The discovery of Sanskrit in 1786 by Sir William Jones is a key event in the History of Linguistics because it initiates the age that extends up to the present time. The era of Modern Linguistics, as it is named, thus covers two centuries that, as will be seen, differ greatly in their linguistic methodology and orientation.

Despite being the age of Comparative and Historical Linguistics, the nineteenth century must be considered a heterogeneous era where three stages can, at least, be identified: the early, the mid-, and the late nineteenth century. The initial phase of the century is, for example, particularly important for being the period when the birth of this new linguistic tradition takes place. The resemblance of Sanskrit to classical languages such as Latin and Greek and also to several Germanic languages brings to light at the end of the eighteenth century two assumptions that, as illustrated in the works by Rasmus Rask (1787–1832), Jacob Grimm (1785–1863), and Franz Bopp (1791–1867), become the hallmark of the linguistic research initiated in the early Romantic age: (a) the existence of a common ancestor language, named Indo-European, for these apparently unconnected languages; (b) and consequently, their possible comparison, as well as, by extension, that of all languages over the world. Notice, in fact, that the first sound law intended to reconstruct the phonological system of Indo-European, known as Grimm's law, and the first linguistic typology established by Wilhelm von Humboldt (1767–1835) according to the morphological pattern the Indo-European languages exhibit, go back to this specific period.

The mid-nineteenth century is, in turn, remarkable because, transferring to the linguistic situation the growing interest of the time in history and the natural sciences, it entails the full establishment of the comparative and historical method of linguistic analysis. As a consequence, the strictly contrastive interest of the previous stage is supplemented in this intermediate phase of the Romantic age, as reflected in its most significant contributions, with an enormous desire to investigate the historical and genetic evolution of languages: August Schleicher's Genealogical Tree Model (1821–1868) is, for instance, designed in consonance with the Darwinian theory to establish the genetic relationships between Indo-European and its descendants, and Johannes Schmidt's Wave Model (1872) is alternatively devised to explain the fact that genetically unconnected languages can and, in fact, do share several linguistic phenomena.

The beliefs pervading linguistic thought up to then are altered late in the century due to the neogrammarian doctrine (1870–1900) defended by scholars such as Hermann Osthoff, Karl Brugmann, and Hermann Paul. The neogrammarians, thus, initiate a new phase in the development of Comparative and Historical Linguistics, which has to be understood, in general, as a reaction against the research carried out previously. Notice, for example, that they reject, among others, Humboldt and Schleicher's idea about the organic nature of language, and that the grammatical laws they postulate do not allow, in contrast to the earlier ones, any kind of irregularity; hence, their need to develop the concept of analogy.

The age of Comparative and Historical Linguistics comes, nevertheless, to its end at the beginning of the twentieth century with the posthumous publication of Ferdinand de Saussure's *Cours de linguistique générale* (1916), which, as deduced from its most basic contents, means a return to Descriptive Linguistics: (a) as a system of interdependent phonological, syntactic, and lexical elements, a language has form and substance; (b) the linguistic competence of a speaker is different from his real linguistic production; hence, the distinction between *langue* and *parole*; and (c) the study of language can be approached either from a diachronic or from a synchronic perspective.

The influence of this work is, in fact, quickly seen in Europe, where conferences, publications, and schools with a clear structural orientation constantly appeared during the first half of the century: among others, the Prague School with Nikolaj Trubetzkoy and Roman Jakobson, the Copenhagen School with Louis Hjelmslev and the London School with John Rupert Firth. Although also pervading the American thought up to the 1950s, the structural linguistic trend takes a slightly different course in the United States owing to the following facts: (a) the existence of unknown American-Indian languages, which arouses an anthropological interest in linguists such as Franz Boas and Edward Sapir; and (b) the influence of behaviorist psychology, which developed around the hypothesis that all kinds of analysis must be based on describable facts, excludes the semantic component of language from the structural theories of outstanding linguists such as Leonard Bloomfield.

The publication of Noam Chomsky's *Syntactic structures* in 1957 finishes, nevertheless, with the

structural theoretical framework and initiates, consequently, a new period in the History of Descriptive Linguistics. The era of Transformational-Generative Grammar, as it is called, signifies a sharp break with the linguistic tradition of the first half of the century both in Europe and America because, having as its principal objective the formulation of a finite set of basic and transformational rules that explain how the native speaker of a language can generate and comprehend all its possible grammatical sentences, it focuses mostly on syntax, and not on phonology or morphology, as structuralism does. Its situation changes, however, from the 1980s onward because, with the increasing interest in meaning awakened in this particular period, the Transformational-Generative model ceases to be the only theoretical framework of the time and instead coexists with several new approaches to the study of language, labeled as functional for regarding, in sharp contrast to the latest version of the Chomskian theory, semantics and pragmatics as the most basic levels of linguistic analysis. In any case, the last modification of Transformational-Generative Grammar nowadays continues to be the leading linguistic theory from among all the formal ones that, for their part, focus mainly on the structural components of language and show, consequently, an almost complete disregard for meaning.

In sum, the age of Modern Linguistics lasts two centuries, which, although heterogeneous in themselves, differ greatly in their linguistic interests: the nineteenth century mainly devoted to the comparison and historico-genetic evolution of languages, and the twentieth century chiefly concerned, in turn, with the description of the linguistic system.

References

Bierwisch, Manfred. 1971. Modern linguistics. Its development, methods and problems. The Hague and Paris: Mouton.

Botha, R.P. 1992. 20th century conceptions of language. Oxford: Blackwell.

Koerner, E.F.K. 1976. Towards a historiography of linguistics. 19th and 20th century paradigms. History of linguistic thought and contemporary linguistics, ed. by Herman Parret. Berlin and New York: Walter de Gruyter.

Marcos Marín, Francisco. 1990 Introducción a la Lingüística: Historia y Modelos. Madrid: Síntesis.

Mohrmann, C., A. Sommerfelt, and J. Whatmough. (eds.) 1961. Trends in European and American linguistics 1930–1960. Utrecht: Spectrum.

Mohrmann, C., F. Norman, and A. Sommerfelt. (eds.) 1963. Trends in modern linguistics. Utrecht: Spectrum.

Mounin, Georges. 1967. Histoire de la linguistique dès origines au XXe siècle. Paris: Presses Universitaires de France.

Mounin, Georges. 1972. La linguistique du XXe siècle. Paris: Presses Universitaires de France.

Newmeyer, Frederick J. 1998 Language form and language function. Cambridge, MA: MIT Press.

Pedersen, Holger. 1931. The discovery of language: linguistic science in the nineteenth century. Bloomington: Indiana University Press; 2nd edition, 1962.

Robins, R.H. 1967. A short history of linguistics, London: Longman.

Waterman, J.T. 1963. Perspectives in linguistics: an account of the background of modern linguistics. Chicago: Chicago University Press.

BEATRIZ RODRÍGUEZ-ARRIZABALAGA

Modification

Two properties semantically motivate the syntactic categories clause and noun phrase, respectively: predication and reference. Clauses involve a predication, that is, they make a judgment about an entity, as in *The weather was changeable*. Noun phrases exhibit the property of reference, whereby linguistic entities make reference to extralinguistic entitites, as in *the green pencil*. Fully referential noun phrases may be both determined and modified. Pronominal noun phrases do not admit either determination or modification (*****that** she* or ***outgoing** he*, for instance, are ungrammatical in Contemporary English, which is indicated by the presence of the asterisk preceding the hypothetical expression); noun phrases whose head is a name instead of a noun undergo determination and modification under certain circumstances, as in ***the big** Jack*.

Determination is a function of the noun phrase that is performed, in English, by grammatical categories (as opposed to lexical categories, which include the noun, the adjective, the verb, and the adverb): the closed classes of the article (*a* and *the*), the demonstrative (*this-these/that-those*), the quantifier (*some*, *much*, etc.; also numerals like *one-first*, etc.) and the possessive adjective. Unlike determination, modification is not exclusive of the noun phrase.

Modification is a function of the noun phrase, the adjectival phrase, the adverbial phrase, and the prepositional phrase, which is realized by the syntactic categories of phrase and clause. Instances of modification in the noun phrase, the adjectival phrase, the adverbial phrase, and the prepositional phrase are, respectively: the **sharp** edge, difficult **to handle**, **incredibly** fast, and **well** below zero. These examples illustrate the contrast between phrasal and clausal modification: the adjectival phrase sharp modifies the nominal head edge, the adverbial phrase incredibly modifies the adverbial head fast, and the adverbial phrase well modifies the prepositional below, whereas the embedded clause to handle modifies the adjectival head difficult. Phrasal and clausal modification give rise to different degrees of phrase elaboration: phrasal modification turns out simple phrases, such as the noun phrase a paltry ten pounds; whereas clausal modification produces complex phrases, as is the case with the adjectival phrase impossible to forget.

In English, the inventory of phrases as modifiers includes the noun phrase, the adjectival phrase, the adverbial phrase, and the prepositional phrase (in other words, all phrasal syntactic categories with the exception of the verb phrase). Noun phrases as modifiers of other noun phrases belong to two classes: they are either nominal compounds (lexicalized or not), as in the **picture** pen, or possesssive modifiers in the genitive, as in **Sally's** promotion. Adjectival phrases modifying nominal heads fall into the descriptive and classifying subtypes: descriptive adjectival phrases focus on a certain property of the noun that functions as a nominal head, as in the **high** wall; classifying adjectival phrases assign a label of type or class to the noun in function of nominal head, as in the **regimental headquarters**. Even though both descriptive and classifying adjectival phrases perform the function of restricting the semantic scope of the modified, descriptive adjectival modifiers in English admit intensification with very (a **very high** tower), whereas classifying adjectival phrases do not (****very** regimental headquarters*). Typically, adverbial phrases modify adjectival heads in adjectival phrases, as in **surprisingly** difficult. Adverbial phrases also function as modifiers of other adverbial phrases, as is the case with **very** deeply. Instances of adverbial phrases as modifiers of noun phrases include adverbial genitives like today's in **today's** top ten. Prepositional phrases modifying nominal heads typically involve possessive postmodification as in the branch **of the tree**. Postmodification by means of prepositional government is also present in noun phrases like the shop **with the dusty window**. An instance of a prepositional phrase modifying an adverbial head is far away **from here**. All the examples given above focus on the

recursive (i.e. repetitive) character of modification: sometimes, modifiers contain heads that are modified themselves, as in the woman with the glittering jewellery, where the prepositional phrase that modifies the nominal head woman governs a noun phrase in which the nominal head jewellery is modified by the adjectival phrase glittering.

Examples like the **sharp** edge, difficult **to handle,** and **incredibly** fast also show the difference between premodification and postmodification in English. The linear order is modifier–modified in the sharp edge and incredibly fast whereas in the adjectival phrase difficult to handle the order qualifies as a modified–modifier. The former linear order is called premodification and the latter is called postmodification. Also of structural import is the question of the syntactic (dis-) continuity of modifiers. When modifier and modified are adjacent, the linear arrangement is described as syntactically continuous, as in **a man who wore a panama hat** called in, where no constituent has been inserted between the relative clause and its antecedent. If modifier and modified are not adjacent, the linear arrangement is termed syntactically discontinuous, as in **a man** called in **who wore a panama hat**, where the verb phase separates the modified (the antecedent) and the modifier (the relative clause). Whereas relative order in the phrase and adjacency represent syntactic phenomena, the phenomenon of double (or multiple) modification is essentially semantic since the structural arrangement of the modifiers with respect to the head reflects the degree of semantic integration of the entity and the properties that are associated with it (typically, in the context of several adjectives modifying a single noun, as in a **small round green** box).

Considering clauses as modifiers, they can be either finite or nonfinite, depending on whether the verbal predicate agrees with the subject or not. Finite clauses in the function of modifier include the relative clause, the nominal modifier clause, and the adjectival modifier clause. Relevant examples of relative clauses modifying nominal heads are the man **who married my sister** (unreduced), and the woman **with the walking-stick,** the manager **promoted by the new directors,** and the girl **riding that bike** (reduced). Nominal modifier clauses can be ilustrated by means of examples like the conclusion **that the plan is feasible**. Adjectival modifier clauses contain, typically, to-infinitive—thus, nonfinite—clauses in adjectival constructions like easy **to misinterpret**.

Morphologically, modifying categories, particularly adjectives, may be either marked or unmarked, that is, variable or invariable. In Contemporary English, modifiers are morphologically unmarked or invariable (with the exception of the genitive in position of premodification -**John's** mistake- and postmodification

-a mistake of *John´s*). Consequently, the tendency is for the modified element to carry the morphological mark (the distinctive feature). In other historical stages of the English language, such as Old English, the modifying adjective showed agreement of gender, number, and case with the nominal head, thus bearing the morphological mark.

References

Aarts, Flor, and Jan Aarts. 1988. English syntactic structures. Functions and categories in sentence analysis. New York and London: Prentice-Hall.

Crystal, D. 1980. A dictionary of linguistics and phonetics. Oxford: Blackwell; 3rd. edition, 1991.

Dik, Simon. 1989. The theory of functional grammar, part 1: the structure of the clause. Dordrecht: Foris.

Dik, Simon. 1997. The theory of functional grammar, part 2: complex and derived constructions, ed. by Kees Hengeveld. Berlin and New York: Mouton de Gruyter.

Downing, Angela, and Philip Locke. 1992. A university course in English grammar. New York and London: Prentice-Hall.

Langacker, Ronald. 1991. Foundations of cognitive grammar II: descriptive application. Stanford: Stanford University Press.

Van Valin, Robert D., and Randy LaPolla. 1997. Syntax: structure, meaning and function. Cambridge: Cambridge University Press.

JAVIER MARTÍN ARISTA

See also **Determiner; Predication; Syntactic Category**

Mohawk and the Iroquoian Languages

Mohawk is a language of the Iroquoian family. The family consists of two major branches: Southern Iroquoian and Northern Iroquoian. Southern Iroquoian is represented by just one language, Cherokee, now spoken primarily in North Carolina and Oklahoma.

Northern Iroquoian has several sub-branches. The first offshoot of Northern Iroquoian developed into Nottoway, Meherrin, and Tuscarora. The Nottoway people were first encountered by Europeans near the Virginia coast in 1650. The language, which disappeared during the mid-nineteenth century, is known through just two wordlists from the early part of that century. The Meherrin people were first encountered in 1650 near the North Carolina coast, but by 1730 they had merged with the Tuscarora. All that remains of their language are two town names. The Tuscarora were first encountered in eastern North Carolina. Early in the eighteenth century, most began to move northward, where their descendants reside today in two locations: near Niagara Falls in eastern New York State and at Six Nations in southern Ontario. Few speakers remain.

The second offshoot of Northern Iroquoian was Huron. The earliest mention of the Huron people is in Champlain's account of his 1615 visit to what is now southern Ontario. The Huron Confederacy, consisting of four tribes, was decimated in 1649 by attacks from the Five Nations Iroquois. Some survivors fled toward Québec City, where their descendants live today at Lorette. The language is no longer spoken there. Others settled with other Iroquoian groups in the area. Some of these groups were subsequently defeated as well, and the survivors, a mixture of Huron, Petun, Erie, and Neutral, moved west to Sandwich, Ontario, and became known as the Wyandot. Many subsequently moved southward, ultimately ending up in Oklahoma. The Wyandot dialect of Huron was last spoken in the mid-twentieth century. French missionaries among the Huron, particularly during the seventeenth and eighteenth centuries, left rich records of the language. Wyandot texts and lexical and grammatical material were recorded in the early twentieth century. Petun, Neutral, Wenro, and Erie are known only through some names.

Five of the remaining Northern Iroquoian groups, the Seneca, Cayuga, Onondaga, Oneida, and Mohawk, formed a strong political alliance known as the League of the Five Nations Iroquois. Their territories stretched essentially from the western edge of modern New York State, where the Senecas were known as the Keepers of the Western Door, to the eastern edge, where the Mohawks were the Keepers of the Eastern Door. Another group to the north, now known as the Laurentian, was not part of the League, nor was a group to the south, the Susquehannock. The Laurentians met Jacques Cartier in 1534 at the mouth of the St. Lawrence River, but they had disappeared from the area by the time of Champlain's arrival in 1603. Vocabulary lists of their languages remain from the time of Cartier's voyage, and it is from them that we have the name *Canada*, a term that persists in the modern Five Nations languages today meaning 'settlement, town'. The Susquehannock to the south, also known as the Andaste, suffered during the

colonial period from European diseases and attacks from colonists and the other Iroquois until the last survivors were murdered in 1763. Their language is known through a wordlist recorded by a Swedish missionary in his journal published in 1696.

The Five Nations languages are all still spoken today primarily by elders, although all communities have language classes. Seneca is spoken in three communities in western New York: Cattaraugus, Allegany, and Tonawanda. Cayuga is spoken at Six Nations in southern Ontario, where many Cayugas fled after the American Revolution. Other Cayugas moved into Oklahoma, where the language was spoken until the late twentieth century. Onondaga is spoken in central New York south of Syracuse and at Six Nations. Some Oneidas remain in New York State, but most moved to southern Ontario near London, and to Wisconsin near Green Bay. There are six main Mohawk communities: Six Nations in Ontario; Tayendinaga near Deseronto in Ontario; Ahkwesahsne with territory in Ontario, Québec, and New York State; Kanehsatake at Oka to the northwest of Montreal; Kahnawake on the southern shore of the St. Lawrence across from Montreal; and Wahta at Gibson in Ontario. There are several thousand speakers of Mohawk, more than any of the other Northern languages. The ages of the youngest speakers vary from community to community, but there are immersion schools in four of them, at Six Nations (Ohsweken), Ahkwesahsne, Kanehsatake, and Kahnawake. A number of speakers of Iroquoian languages reside outside of these communities as well.

Although the languages are not mutually intelligible, their basic structures are similar. They are illustrated here with examples from Mohawk.

The consonant inventory is small: *t, k, kʷ, ts, s, n, r, y, w, h, ʔ.* There is a notable absence of labials (such as *p, b,* or *m*). The vowel inventory is similarly small: *i, e, a, o, ũ, ʌ̃.* The Mohawk communities have agreed on a practical orthography, which represents the consonants as *t, k, kw, ts, s, n, r, i, w, h, and '* (with *i* for the glide [j] and apostrophe for glottal stop). The vowels are spelled *i, e, a, o, on,* and *en,* with nasalization indicated by a following *n.* Stress originally occurred on the penultimate syllable of a word, although the pattern can be obscured by the addition of vowels within the word. The stressed syllable carries distinctive tone (pitch). An acute accent marks high or rising tone (*ó*), and a grave accent marks falling tone (*ò*). The falling tone appears where a stressed vowel immediately preceded a laryngeal (such as *h*), as in **iahratóhrok',* which became *iahatò:roke'* 'he climbed under there'. If a syllable ended in a laryngeal, the laryngeal was lost,

leaving just vowel length. Open, stressed syllables are lengthened, as in *kí:ken* 'this'. Length is marked orthographically with a colon:.

Nouns are based on a noun stem, which may be a noun root or a nominalized verb stem. Morphological nouns begin with a noun prefix. In unpossessed nouns, the prefix encodes the gender of the referent, like the Neuter *o-* in <u>*o*</u>*-tsíhkw-a'* 'fist, knot, knob, puck'. If the noun is possessed, the prefix encodes the person, number, and gender of the possessor, as in <u>*akw*</u>*-atsíhkwa'* '<u>my</u> puck'. Different sets of possessive prefixes are used for Alienable and Inalienable possession. The prefix *akw-* 'my' in 'my puck' indicates Alienable possession (meaning that the possessed is not an integral part of the possessor). The prefix *k-* 'my' in <u>*k*</u>*-atsihkwà:ke* '<u>my</u> fist' indicates Inalienable possession (meaning that the possessed is an integral part of the possessor). Inalienable possessions include most body parts, but not hair or internal organs such as the heart or stomach. Separated body parts such as eyelashes or fingernails are Alienable. Terms for most kinsmen do not describe possession, but rather the relationship, such as *rake'níha* 'my father', literally 'he is father to me', or the reciprocal *atiara'sè:'a* 'my cousin', literally 'we two are cousins to each other'.

Morphological nouns end in a noun suffix. The most common is *-a'* as in *otsíhkw-*<u>*a'*</u> 'knob'. Various grammatical endings can be added to words serving as nominals, as in *Kahnawa'kehronon'kénha'* 'former Kahnawake residents': *ka-hnaw-a'=ke=hronon'= kenha'* NEUTER - rapids - NOUN.SUFFIX = LOCATIVE = RESIDENTIAL = DECESSIVE.

Verbs are potentially the most complex words and by far the most frequent in speech. All verbs contain a pronominal prefix, a verb stem, and, apart from commands, an aspect suffix. This structure can be seen, for example, in *katerohrókha'* 'I watch': *k-aterohrok-ha'* I-watch-HABITUAL. The pronominal prefix represents the core arguments of the clause, that is, the one or two major persons or objects involved. These specify person (first, inclusive, exclusive, second, or third), number (singular, dual, or plural), and in third person, gender (masculine, neuter-zoic, or feminine-indefinite). The grammatical roles of the arguments are specified as well, but not in terms of subjects and objects, but rather as grammatical Agents (typically those performing actions and controlling situations) and grammatical Patients (typically those affected by the situation but not in control). The categories are semantically based, but they are crystallized in the lexicon and the grammar, so that speakers have no choices about degrees of agency as they speak. Verb forms are simply learned with the appropriate prefixes. The

intransitive verb *eniakwaterohrókha'* '<u>we</u>'ll go watch it' contains the Agent pronominal prefix *iakw-* 'we all'. The intransitive verb *ionkwén:ten* '<u>we</u> are poor' has the Patient pronominal prefix *ionkw-* 'we'. The Agent and Patient components of transitive pronominal prefixes are often fused. The transitive verb *shakwá:iatskwe'* '<u>we</u> called <u>him</u>' has the transitive pronominal prefix *shakwa-* 'we/him'. The transitive verb *enionkhiia'tahserón:ni'* '<u>she</u> will dress <u>us</u> up' contains the pronominal prefix *ionkhi-* 'she/us'.

The verb stem may itself be complex. This stem *aterohrok* 'watch' contains a Middle prefix *-ate-*. Verb stems may also contain an incorporated noun stem, like *-nahskw-* 'domestic animal' in *ranahskwiióhne'* 'he was a beautiful animal': *ra-<u>nahskw</u>-iio-hne'* MASCULINE.AGENT-<u>animal</u>-be.beautiful.STATIVE-PAST ('he was beautiful in the way domestic animals are'). Incorporated nouns qualify the meaning of the verb. Nouns are incorporated both to create single words for recurring concepts, as above, and to manipulate the flow of information. When speakers wish to direct special attention to an object, they generally designate it with a separate, independent noun. If the object is an established part of the scene, or incidental to the point at hand, it may be backgrounded by incorporation.

Verbs may contain various additional prefixes and suffixes. Among the prefixes are a Contrastive for unexpected situations, a Coincident for similarity or simultaneity ('the same, when'), a Partitive 'so' that appears in a variety of syntactic constructions, a Negative 'not', a Translocative 'thither', a Factual typically used for past events, a Duplicative that indicates various kinds of 'two-ness', including repetition of an event or a shift in position or state, a Future tense 'will', an Optative 'might, should, would, could', a Cislocative 'hither, there', and a Repetitive 'again, back'.

Among the suffixes are an Inchoative 'become', several Causatives ('cause'), Instrumental Applicatives ('do with …'), Benefactive Applicatives ('do for'), Reversives ('un-'), Distributives ('here and there'), and Purposives ('go in order to …'). All verbs except commands contain one of the three basic aspectual suffixes: Habitual, Punctual, or Stative. The Habitual is used for recurring events and, with some verbs, for ongoing activities. The Punctual is a Perfective, used for events viewed as wholes. The Stative is used for states. With some verbs, it is also used for activities in progress, and with some verbs it is also used as a Perfect. The Habitual and Stative may be followed by a postaspectual suffix: Past, Continuative, or Progressive.

Because all verbs contain pronominal reference to their core arguments, they can serve as complete sentences in themselves: *wa'onkwanahskwaién:ta'ne'* 'we got a pet' (*wa'-onkwa-nahskw-a-ient-a'-n'* FACTUAL-1.PLURAL.PATIENT-domestic.animal-STEM.JOINER-have-INCHOATIVE-PUNCTUAL). The same verb can be part of a larger sentence: *wa'onkwanahskwaién:ta'ne' è:rhar* 'we got a dog'. Verbs can serve other syntactic functions as well. They can function syntactically as nominals, much like nouns. Some are lexicalized as nominals, so that speakers understand them first as names for entities, such as *tewa'á:raton* 'lacrosse stick' (*te-w-a'ar-a-t-on* CISLOCATIVE-NEUTER-net-STEM.JOINER-be.in-STATIVE 'it has a net in it'). Some are used alternately as predicates or nominals, like *tahontsihkwà:'eks* 'they hit the puck, they play lacrosse' or 'lacrosse players' (*ta-hon-tsihkw-a-'ek-s* CISLOCATIVE-MASCULINE.PLURAL.AGENT-puck-STEM.JOINER-hit-HABITUAL). Full verbs are also used for many functions served by adjectives and adverbials in other languages.

Because the verb provides a full grammatical skeleton of the sentence, word order is used for purely pragmatic purposes. Nouns are strikingly rare in connected speech; sentences most often consist of just a verb and various particles. When independent nominals are present in a clause, all possible constituent orders can occur, although not all would be pragmatically felicitous. After various orienting particles, the most important elements tend to occur early in the clause, with successively more predictable and peripheral information expressed later.

Some of these structures can be seen in the excerpt below from an anecdote told by Watshenní:ne' Sawyer of Kahnawake. (Abbreviations include M for MASCULINE, N for NEUTER, PL for PLURAL, PRT for PARTITIVE, SJ for STEM JOINER, and ST for STATIVE.)

Nè:ne   ó:nen
it is   now
'Now then

wa'-onkwa-nahskw-a-ién:t-a'-ne'	*è:rhar.*
FACTUAL-1.PL.PATIENT-ANIMAL-SJ-LIE-	
INCHOATIVE-PUNCTUAL	dog

we acquired a dog.

Ra-nahskw-iió-hne'	*ken'=k*	*nì-:r-a*	*è:rhar.*
M.AGENT-animal-	little=just	PRT-M.SG.	dog
be.nice.ST-PAST		AGENT-	
		be.a.size	

He was a beautiful little dog.

Butch   ronwá:-iat-s-kwe'
        3.PL/M.SG-call-HABITUAL-PAST
His name was Butch.

Toka'   kí:ken   Ka-hnaw-a'=ke=hró:non'

and   this   NEUTER-rapids-NOUN.SUFFIX=
                 place=RESIDENTIAL
And folks in Kahnawake,

thi   ionkw-én:ten   shen's   ki'   wáhi'
that   1.PL.PATIENT-be.poor   formerly just   TAG
we were poor then, you know,

énska=k   ki'   nà:'a
one=only   just   I guess
I guess we only had one

wa'-t-hon-tsihkw-a-'ék-st-ha'
FACTUAL-DUPLICATIVE-M.PL.AGENT-puck-SJ-hit-
INSTRUMENTAL.APPLICATIVE-HABITUAL
ball ('they hit the puck with it', i.e. 'they played
lacrosse with it').

Ahkwesáhs=ne   n-en-t-hón:n-e-'
Ahkwesahs=place   PARTITIVE-FUTURE-CISLOCATIVE-
                 M.PL.AGENT-go-PUNCTUAL
They would come from Ahkwesahsne,

ta-hon-tsihkw-à:-'ek-s
CISLOCATIVE-M.PL.AGENT-puck-hit-HABITUAL
the lacrosse players

aw-ent-a-tokenhti-'=ke   io-t-ohetst-on n=entie'.
N-day-SJ-be.holy-NOMINALIZER N.PATIENT-   the=noon
=place   MIDDLE-pass-ST
Sunday   afternoon.

Sok   en-ionkhi-ia't-a-hserón:ni-'   istèn:'a
then   FUTURE-INDEFINITE/   mother
        1.PL-body-SJ-prepare-P
So then my mother would dress us up

sok   i-en-ionkhi-ia't-énhaw-e'
then   TRANSLOCATIVE-FUTURE-INDEFINITE/1.PL-body-
       take-PUNCTUAL
and then she'd take us over there

en-iakw-ate-rohrók-ha-'   kí:ken.
FUTURE-1.PL.EXCLUSIVE.AGENT-   this
MIDDLE-watch-PURPOSIVE-PUNCTUAL
to watch the game.'

References

Chafe, Wallace. 1976. The Caddoan, Iroquoian, and Siouan languages. The Hague: Mouton.

Froman, Frances, Alfred Keye, Lottie Keye, and Carrie Dyck. 2002. English-Cayuga Cayuga-English dictionary. Toronto: University of Toronto.

Michelson, Karin, and Mercy Doxtator. 2002. Oneida-English English-Oneida dictionary. Toronto: University of Toronto.

Mithun, Marianne. 1979. Iroquoian. The languages of native America: historical and comparative assessment, ed. by Lyle Campbell and Marianne Mithun. Austin and London: University of Texas.

Mithun, Marianne. 1999 and 2001. The languages of native North America. Cambridge, UK: Cambridge University.

Rudes, Blair A. 1999. Tuscarora-English English-Tuscarora dictionary. Toronto: University of Toronto.

Trigger, Bruce G. (ed.) 1978. Northeast: handbook of North American Indians, Vol. 15. Washington: Smithsonian Institution.

Woodbury, Hanni. 2003. Onondaga-English English-Onondaga dictionary. Toronto: University of Toronto.

MARIANNE MITHUN

See also **Noun Incorporation**

Mongolian

The language known as Mongolian, one of several languages belonging to the Mongolian subgroup of the Altaic languages, is based on the Khalkha language. It is spoken in northeastern Asia by approximately 2.5 million people in the Mongolian People's Republic, commonly known as Outer Mongolia, where it is the official language. In addition, about 2 million people in the Inner Mongolian Autonomous Region of the People's Republic of China and up to half a million people in other areas of northern China speak Khalkha Mongolian. Other languages that belong to the Mongolian subgroup of the Altaic family include Oirat in Outer Mongolia and China, Kalmyk along the lower Volga River in the Kalmykia Autonomous Soviet Socialist Republic and in parts of China, Mogul in Afghanistan, Buryat in the Buryatia Autonomous Republic and in parts of China, Dagur in Heilongjiang and the Inner Mongolian Autonomous Region of China, Santa in the Gansu region of China, and Monguor and Bao'an, both in the Qinghai region of China.

As an Altaic language, Mongolian shares several characteristics with other languages in this family. Some of these include agglutination, the use of postpositions instead of prepositions, vowel harmony, the placement of modifiers before what they modify, the absence of a relative pronoun, the absence of a verb meaning 'to have', the absence of grammatical gender, and the absence of articles. Despite these

similarities, some linguists do not accept the thesis of a common 'Altaic' origin. They claim that they may be due to sociolinguistic contact and mutual interaction over millennia, which would make the similarities typological (due to mutual contact) instead of genetic (due to a common origin). This is complicated by the obscurity of the origin of the 'Altaic' -speaking peoples, the extreme nomadism of these peoples, and, simply, linguistic change.

As an agglutinating language, Mongolian forms words by adding morphemes (meaningful elements that cannot be split into smaller meaningful elements) to a base, or root, word with each morpheme corresponding to a single lexical meaning or grammatical function. An illustration of Mongolian agglutination is the word *bari-ld-aa-tʃi-d-t* 'to the wrestlers', which is composed of the following parts: *bari-* 'to seize' + *-ld-* 'each other' + *-aa-* NOUN-FORMING SUFFIX + *-tʃi* 'an occupation' + *-d-* PLURAL + *-t* LOCATIVE. Mongolian also has a number of cases (these manifest themselves as suffixes that attach to nouns); these cases indicate the function of the word in a sentence. The following is a chart of the declension of the words *mal* 'cattle' and *ger* 'yurt'.

Case	'cattle'	'yurt'
Nominative (indicates the subject)	*mal*	*ger*
Genitive (indicates possession 'of')	*mal-iin*	*ger-iin*
Dative-Locative (indicates the indirect object or location 'in'/'at')	*mal-d*	*ger-t*
Accusative (indicates the direct object)	*mal-iig*	*ger-iig*
Ablative (indicates direction away from 'from')	*mal-aas*	*ger-ees*
Instrumental (indicates means or agency 'with'/'by')	*mal-aar*	*ger-eer*
Comitative (indicates accompaniment 'with')	*mal-taj*	*ger-tej*

As illustrated by this chart, many suffixes in Mongolian correspond to prepositions in English. For example, the Mongolian word *mal-iin* would be translated into English as 'of the cattle'. It is possible to combine more than one of the above suffixes, resulting in constructions such as *nœxœr-* 'friend' + *-iin* 'of' GENITIVE + *-d* 'at' LOCATIVE = *nœxr-iin-d* 'at my friend's house'. Mongolian also has some postpositions (these are like prepositions except that they occur after nouns), which follow nouns in the nominative, genitive, ablative, or comitative cases:

> with nominative: *dotor* 'in', e.g. *tasalgaan dotor* 'in the room';
> with genitive: *tuxay* 'about', e.g. *nom-iin tuxay* 'about the book'; etc.

One of the striking features of Mongolian (and Altaic languages in general) is vowel harmony. Vowel harmony occurs within a word, including the suffixes that have been added. When vowels harmonize, all of the vowels within a word agree—they must all either be front (/e/, /y/, /œ/) or back (/a/, /o/, /u/), with /i/ being a neutral vowel. Because of this, many suffixes have two versions: one with a front vowel and one with a back vowel. Vowel harmony is demonstrated in the ablative, instrumental, and comitative cases in the above chart; because of vowel harmony, the form **ger-aas* is ungrammatical because the suffix does not have a front vowel as the stem word. Vowel harmony rules, however, do not apply in four instances: compound words, where each member retains its own vowels (*tserendulmaa* 'Tserendulmaa' (a woman's name)); certain suffixes (*oj-gwe* 'treeless, bare of forests'); foreign names (*germaani* 'Germany'); and recent borrowings (*pjoneer* 'pioneer' (Boy Scouts)).

The basic word order of Mongolian is subject–object–verb. Modifiers (adjectives, adverbs, determiners, etc.) generally precede the word that they modify and do not show agreement, e.g. no matter what case (nominative, genitive, dative, etc.) or number (singular or plural) a noun is, the adjective will stay the same. Although Mongolian has various plural markers, e.g. *nom* 'book', *nom-uud* 'books', *zaluu* 'youth', *zaluu-tsuud* 'youths', a noun does not take a plural suffix when modified by a number, e.g. *negen nom* 'one book', *xoyor nom* 'two books', *gurav nom* 'three books'. Because the number already indicates whether there is one or more, a plural marker would be redundant. The following sentence illustrates many of the features just described.

> *bi xojor sajn mori-toj   bajn*
> I two good horse-COMITATIVE exist
> 'I have two good horses.'

The adjective *sajn* 'good' shows up before the noun *mori-toj* 'horse', which is in the comitative case, but it does not show agreement. If it were to show agreement, one would expect to find the comitative suffix added to *sajn*. Also, the noun *mori-toj* 'horse' is modified by the number *xojor* 'two', but it does not have a plural marker. The equivalent sentence with the plural marker would be

> *bi sajn mori-d-toj   bajn*
> I good horse-PLURAL-COMITATIVE exist
> 'I have good horses.'

This sentence also shows how Mongolian expresses possession without a verb 'to have': the verb meaning 'to be/to exist' is used with the possessor in the nominative case and the possessed in the comitative case. The literal translation would be 'I am with good horses.'

For comparative structures ('NOUN is ADJECTIVE-er than NOUN'), Mongolian puts the

first noun in the nominative case, the second noun ('than NOUN') in the ablative case, and then the adjective. The sentence 'Today is warmer than yesterday' would be translated *ænœædœr œtʃigdr-œœs dulaan* (today yesterday-ABLATIVE warm).

Mongolian differs from other Altaic languages in verb endings: Mongolian verbs are not marked for person. Rather, verbs only inflect for tense (past vs. nonpast) and for aspect (perfective vs. imperfective); the bare verb stem serves as an imperative (a command). There are only four basic verb endings:

	Past	Nonpast
Perfective	*-v*	*-laa*
Imperfective	*-dʒee*	*-na*

In addition to these suffixes, one can add *–aarai* to a bare stem to form a polite request: *yav* 'go!' becomes *sajn yawaarai* 'please go well!'

A verb with one of the basic verbal endings cannot be negated. The negation of a verb is achieved by attaching the negative suffix, *-gwe,* to a noun form of a verb. There are three noun-forming endings for verbs: *-san* for past tense, *-x* for nonpast, and *-dag* for customary or repeated actions. Hence, for instance, the sentences *ten ire-v* 'he came' and *bi mede-n* 'I know' would have the following negative counterparts: *ten ir-sen-gwe* 'he did not come' and *bi mede-x-gwe* 'I do not know,' not **ten ire-v-gwe* or **bi mede-n-gwe.* In English, the closest literal translation of these phrases would be 'without his having come' and 'without my knowing', respectively, both gerund structures. However, in order to negate a command, one need only add the negative word *bitgii,* as in *bitgii yav* 'Don't go.' The verb *bajn* 'to be/to exist' is also not negated like other verbs; it has its own negative counterpart, *biʃ* 'not to be/not to exist'— *ter tend biʃ* 'he is not there'.

Mongolian does not have relative pronouns. In order to form a relative clause, verbs transform into verbal nouns and are then used as attributive adjectives to modify nouns. For instance, the relative construction 'the work which we did' would be rendered *bidnii xii-sen adʒil* ('we do-PAST work'). The verb *xiix* 'to do' is turned into a noun with the addition of the past-tense morpheme *–san* (in this case, *–sen* due to vowel harmony). Then it, along with the other part(s) of the clause (in this instance, there is only a subject pronoun, *bidnii* 'we'), is placed before the noun that this modifies, *adʒil* 'work'.

The history of the Mongolian language is usually divided into three periods: the ancient period, the middle period, and the contemporary period. The ancient period lasted until about the twelfth century CE. Although texts written in ancient Mongolian have not been found, evidence exists in Chinese texts wherein a few ancient Mongolian words are attested and in modern Tungus languages, whose ancestors borrowed words from ancient Mongolian. Some characteristics of ancient Mongolian include the preservation of the consonants [p], [f], and [h] in initial position; the division of all vowels and consonants into front and back; the preservation of [γ], [g], [b], and [w] in intervocalic (between vowels) position; and the presence of grammatical categories. The middle period lasted from around the twelfth century to the sixteenth century CE. The language of this period is known as Classical Mongolian, and the Mongolian script comes into use at this time. Distinguishing traits of middle, or Classical, Mongolian include the disappearance of [γ], [g], [b] and [w] between vowels (this leads to the development of long vowels in Mongolian); the loss of [h] in initial position in some words; and the partial loss of grammatical categories. The contemporary period, which begins around the sixteenth century CE, follows middle Mongolian and continues to the present. This period is characterized by the loss of [p], [f], and [h] in initial positions; the distinction between short and long vowels; and the absence of grammatical categories.

The Mongols have used several different systems to write their language. The *Secret History of the Mongols,* a major literary work that dates from 1240 CE and marks the beginning of the Middle Mongolian period, was written in Chinese characters. The classical Mongolian script was in use by the thirteenth century. Most scholars believe that the Mongols borrowed this alphabet from the Uigurs who had adopted it from a version of the Syriac alphabet. The letters of this Mongolian alphabet have initial, medial, and final forms (the shape of the letter changes depending on whether it is the first letter, a middle letter, or the final letter of a word), and it was originally written horizontally from right to left. Later, this changed to a vertical, left-to-right format, perhaps under the influence of Chinese. The Mongolian script was used in the Mongolian People's Republic until 1941 when a new Cyrillic-based alphabet was adopted. There were two reasons stated in official documents for this adoption: first, a big gap between the written language (using the traditional Mongolian script) and the spoken language had developed; and second, the Mongolian script was not suitable for assimilating foreign words. Despite this adoption, the Mongolian script continues to be used by Mongolian speakers in China and in private correspondences by older people in Outer Mongolia.

References

Campbell, George L. 1991. Compendium of the world's languages London and New York: Routledge; 2nd edition, 2000.

Deny, J. 1952. Langues Turques, Langues Mongoles et Langues Toungouzes In Meillet and Cohen.

Meillet, A., and Marcel Cohen (eds.) 1952. Les Langues du Monde. Paris: Center national de la recherche scientifique.

Poppe, Nicholas. 1965. Introduction to Altaic linguistics. Wiesbaden: Otto Harrassowitz.

Poppe, Nicholas. 1970. Mongolian language handbook, Washington: Center for Applied Linguistics.

Sinor, Denis. 1952. Langues Mongoles, In Meillet and Cohen.

Troxel, D.A. 1953. Mongolian vocabulary. Washington: United States Government Printing Office.

ROBERT A. CLOUTIER

See also **Altaic; Vowel Harmony**

Montague, Richard

Richard Montague is one of the most influential philosophers of the past century and can be deservedly considered the founder of contemporary formal semantics. As an undergraduate at the University of California at Berkeley (1948–1950), he was attracted to a number of disciplines, but particularly to mathematics, philosophy, and Semitic languages, all of which he pursued very rapidly to an advanced level. The most important single influence on the direction of his work was the logician Alfred Tarski. While still a graduate student at Berkeley, Montague had already acquired considerable national and even international reputation. Between 1954 and his formal dissertation defense in 1957, he authored six, and coauthored another four, significant papers in mathematical logic, including researches on Boolean algebras, proof theory, model theory, and axiomatic set theory. In 1955, he joined the faculty of the University of California at Los Angeles and expanded his interests into other areas in the field of logic such as abstract recursion theory, predicate logic, and the model theory of higher-order logics. In the early 1960s, he increasingly focused on the application of sophisticated logical methods to traditional problem areas in philosophy: determinism (in "Deterministic theories," published in 1962); the concept of scientific explanation; the so-called hangman or surprise examination paradox ("A paradox regained," 1960, with David Kaplan); conditional or derived obligation; events ("On the nature of certain philosophical entities," 1960); and indirect discourse ("That," 1959, with Donald Kalish).

From the mid-1960s until his untimely death, Montague embarked on an ambitious program to apply some of the precise tools of mathematical logic and model theory to the analysis of natural language. This was a concern that started to emerge in his introductory logic text, *Logic: techniques of formal reasoning* (1954, with Donald Kalish). The main advantage of this mathematical approach was achieving an accurate, adequate, and philosophically satisfactory scientific account of natural language. The initial words of "Universal grammar" (1970a) mark a clear declaration of intentions:

> "There is in my opinion no important theoretical difference between natural languages and the artificial languages of logicians; indeed I consider it possible to comprehend the syntax and semantics of both kinds of languages within a single natural and mathematically precise theory."

In this paper, he states that a grammar consists of a syntactic algebra and a semantic algebra that are subject to a homomorphism condition: there is a homomorphism mapping elements of the syntactic algebra onto elements of the semantic algebra. This requirement elegantly captures Frege's Compositionality Principle in a formal setting: the meaning of a complex expression will be a function of the meaning of its parts and the operation or operations combining them.

Montague also introduced the method of fragments. Unlike in generative grammar, where structural descriptions and rules are proposed for a single construction or constructions but the rest of the grammar is left undefined (or assumed to be generated by standard mechanisms), Montague presented a formal and precise treatment of several "fragments" or subsets of expressions of English. The monostratal syntactic formalism he proposed also departed from the standard multistratal proposals preferred within generative-transformational approaches. In the latter, syntactic derivations are based on a set of base rules and additional transformations that change the linear arrangements of terminals, respectively deriving deep and surface structure representations. Montague's syntax is modeled on the recursive bottom-up rules preferred by logicians and categorial grammarians. These rules derive well-formed expressions of different categories, which in turn are assigned an interpretation.

There are two possible methods for providing a semantic interpretation for a fragment of English. The first one, illustrated in "English as a formal language" (1970b), is the method of direct interpretation. Expressions of a language are directly assigned denotations in a model. The second method (the indirect method) is developed in "The proper treatment of quantification in ordinary English" (1973). It consists of translating the expressions of English into a formalized logical language (the language of typed intensional higher-order logic) on a first step and providing an interpretation for this language. Although from a technical point of view, this intermediate step should be dispensable, in practice using a formalized representation language has several advantages. First, it makes possible a systematic analysis of semantic ambiguity. One or more disambiguated formulae are assigned to each ambiguous English sentence. Second, translating English into a rigidified uniform language increases perspicuity and the ability to represent the logical properties of natural language in different realms (tense, modality, quantification) using well-known precise formal devices. Finally, Montague always maintained a nonpsychological objectivistic stance, and considered the syntax, semantics, and pragmatics of natural languages as branches of mathematics on par with geometry or number theory. Nevertheless, in the following decades several connections have been drawn with some of the psychological concerns of linguistic theories, mostly in that the intermediate semantic representation language can be viewed as a characterization of the speakers' semantic competence (or at least of some of its properties).

Numerous innovative semantic proposals by Montague have become standards in the field. First, he treated noun phrases as uniformly denoting (generalized) quantifiers, against the received view in logic that considers proper names as constants and existential, and universals as quantifiers. Second, Montague used lambda-abstraction and the lambda-operator to represent the compositional combination of expressions as functions and arguments. Finally, he generalized the Frege/Carnap distinction between sense/intension and reference/extension to a full intensional semantics. The intension of an expression is a function from possible worlds to its extension. Intensionality is pervasive in natural language, not only in opaque contexts such as the complements of verbs like *believe* but also in the transitive complements of verbs such as *seek*, or in modifiers such as *presumed*, *alleged,* or *former*.

Montague's most significant articles with applications in linguistics and philosophy were posthumously compiled by Richmond Thomason in the volume *Formal philosophy* (Yale University Press,

1974). The groundbreaking nature of Montague's approach can only be properly understood in the context of a comparison between the linguistic landscape of that time and the field of formal semantics nowadays. This discipline has undergone radical changes that are almost entirely attributed to Montague's legacy, namely to the combination of the ideas and analysis contained in his most influential papers and the developments by his disciples and followers, within the enterprise commonly known as Montague Grammar.

Biography

Richard Montague was born in Stockton, California, on September 20, 1930. He received his B.A. from University of California, Berkeley in 1950 and Ph.D. for a dissertation on mathematical logic, tutored by Alfred Tarski, (University of California, Berkeley) in 1957. During his graduate years, he also worked with Professors W.J. Fischel in Arabic, and Paul Marhenke and Benson Mates in philosophy; between 1950 and 1953, he held the Howison Fellowship in Philosophy, and for two succeeding years he was a teaching assistant in mathematics. In the spring semester of 1955, he joined the faculty of the University of California, Los Angeles, as Acting Instructor in Philosophy, and in the following years he advanced very quickly through the academic ranks. He served as a member of the United States National Committee for the International Union of the History and Philosophy of Science and, for the years 1966 and 1967, as Chairman of the national Subcommittee for Logic, Methodology, and Philosophy of Science. He also served as Secretary of the Association for Symbolic Logic from 1966 until his death and as a consulting editor for the *Journal of Symbolic Logic* since 1958. He died on March 7, 1971, in his home in Los Angeles, at the hands of persons still unknown at the time of writing

References

van Benthem, Johan, and Alice ter Meulen (eds.) 1997. The handbook of logic and language. Cambridge, MA: MIT Press.

Dowty, David, Robert Wall, and Stanley Peters. 1981. Introduction to Montague semantics. Dordrecht: D. Reidel Publishing Company.

Gamut, L.T.F. 1991. Logic, language and meaning. Chicago: University of Chicago Press.

Gutiérrez-Rexach, Javier (ed.) 2003. Semantics: critical concepts in linguistics. London/New York: Routledge Publishing Company.

Lewis, David. 1972. General semantics. Semantics of natural language, ed. by Donald Davidson and Gilbert Harman, 169–218. Dordrecht: D. Reidel Publishing Company.

Montague, Richard. 1970. English as a formal language. Linguaggi nella Società e nella Tecnica, ed. by B. Visentini et al., 189—224. Milan: Edizioni di Comunità.

———. 1970. Universal grammar. Theoria 36. 373–98.

———. 1973. The proper treatment of quantification in ordinary English. Approaches to natural languages, ed. by J. Hintikka, J. Moravcsik, and P. Suppes, 221–47. Dordrecht: D. Reidel Publishing Company.

———. 1974. Formal philosophy. Selected papers of Richard Montague, edited and with an introduction by Richmond Thomason. New Haven: Yale University Press.

Partee, Barbara, 1973. Some transformational extensions of Montague grammar. Journal of Philosophical Logic 2. 509–34.

Partee, Barbara (ed.) 1976. Montague grammar. New York: Academic Press.

JAVIER GUTIÉRREZ-REXACH

Mood

The grammatical meaning of *mood* (often also referred to as *modality* or by the adjective *modal*) is a complex one, but one that we can break down into two main characteristics. In its narrow meaning, it denotes a particular aspect of verbal morphology, usually indicated by an inflectional marking on a verb form, tied in closely with tense and aspect relations. But in the larger context of sentences ("form") and meaning ("function"), it refers to the grammatical category expressing the type of sentence. To make matters more complicated, both types are sensitive to the specific modal function of individual elements, in addition to both the narrow and wider understanding of mood.

In traditional grammar, mood expressed in verbal morphology commonly distinguishes the indicative (*John plays soccer.*), the subjunctive (*John would play soccer.*), the imperative (*Play soccer, John!*), and (in Classical Greek or Hungarian) the optative (roughly, *I wish John played soccer.*). These distinctions can be found in many European languages. Other languages express the same moods by the use of particles (e.g. Ngiyambaa, a language from Australia or Jacaltec from Guatemala), and also clitics (such as Luiseño from California). This state of affairs warrants a wider conception of mood already than simply referring to verbal paradigms. Moreover, even those languages that express mood with verbs can be further differentiated. The classical languages (Classical Greek, Latin) and the modern Romance languages beyond French (e.g. Spanish, Italian) have particular verbal paradigms, i.e. inflectional endings on the main verb. The North European languages (such as French, English, or German) use modal verbs, but German and French, for example, can also express moods through verbal inflection, namely the subjunctive.

From a European point of view, the modal verbs are the clearest example of the narrow conception of mood. The traditional modal verbs of English are *may, can, must, ought (to), will*, *shall*, and to some extent *need* and *dare* (including *might, could, would, should*). These have counterparts in German (e.g. *mögen* 'may', *wollen* 'want'*, könn, 'can*') and in French (such as *pouvoir* 'can', *devoir* 'must', *vouloir* 'want'). In this sense, there is an obvious interplay of mood with *tense*, the grammatical distinctions of time made, and *aspect*, relating to the internal temporal structure of a situation (see also the paragraph on epistemic vs. deontic modality).

It must be said, however, that from a typological perspective, this characterization of mood is less than clear: not all mood forms are used equally in all languages—neither with respect to type of expression (e.g. verbal morphology vs. separate expressions) nor variety of expression; one extreme in the latter category is Tuyuca, a language spoken in Brazil and Colombia, where every sentence is modal: Tuyuca has five evidential modals (see below) that characterize even every declarative sentence in a specific modal form. The English sentence *He played soccer*, for example, can be rendered by *díiga apé-wi* 'He played soccer (I saw him play)', *díiga apé-ti* 'I heard the game and him, but I didn't see it or him', or *díiga apé-yi* 'I have seen evidence that he played: his distinctive shoe print on the playing fields. But I didn't see him play', for example.

And comparing German and English, for a moment, we can see another factor of the interplay of different types of modality. The German equivalent of the English sentence *He should have come home* is *Er hätte nach Hause kommen sollen*, where the first boldfaced word is the subjunctive form of the verb 'to have', while the second one is the modal form for 'should', both expressed by one modal form in English (see also the paragraph on epistemic vs. deontic modality below). On the other hand, languages like French use the subjunctive often for purely morphosyntactic reasons, rather than semantic ones, namely when embedded under particular verbs.

Apart from verbal mood classes, mood also expresses the speaker's attitude. Among the classifications of mood distinctions that have been proposed in the literature, the one between *epistemic* and *deontic* use is the most common one. Epistemic modality is the area of mood concerned with knowledge and belief (incl. 'possibility', 'probability', 'certainty'), and it may also express the speaker's degree of commitment to what he says (see also evidential mood below). The deontic use of mood or modality expresses 'permission', 'obligation', and 'prohibition'. For example, *John may play soccer tomorrow* is ambiguous between the possibility of John playing soccer the next day (the epistemic meaning viz. 'possibility') and the 'permission' given to John to play soccer (by the speaker or someone else; the deontic use).

Another proposal considers a three-way distinction of mood between *illocutionary force* (the 'communicative intention of an utterance'; see below), *status* ('degree or kind of reality'), and *modality* (in the sense of declarative, interrogative, etc.). *Evidential* and *evaluative* distinctions are also sometimes regarded as mood distinctions. We have already seen an instance of the former above, indicating the kind of evidence a speaker has for what he says (e.g. direct vs. indirect evidence). Evaluatives, on the other hand, can be defined as attitudes toward known facts, such as Hixkaryana (a Carib language spoken in Northern Brazil), where *nomokyaha hampɨnɨ* 'He's coming—be warned!' vs. *nomokyatxow hampe* 'They are coming! I don't believe it!'.

The larger picture that mood or modality expresses is the *illocutionary force* of an utterance, which may be that of a request for action or information, or an order, a warning, a promise, etc. Here, a strict characterization is even more difficult than with the narrower meanings of mood, because the illocutionary force of an utterance is in general independent of its grammatical form, or *sentence type*. In this connection, we often speak of sentence type referring to the form of a sentence (the above-mentioned moods 'declarative', 'interrogative', etc.), where it is clear that not every interrogative asks for information, for example (viz. *Could you pass the salt, please?*, which requests an action, akin to an imperative or order).

In sum, mood in its narrow sense denotes a particular distinction in the (often verbal) paradigm of languages, either by encoding inflectional morphology, modal verbs, particles or clitics expressing a particular mood. In the broader context, modality relates to the speaker's attitude (e.g. epistemic vs. deontic modality) and the function of a sentence uttered (viz. illocutionary force). There is a rich literature that can be found on each of these topics, both from the point of view of a single language as well as typologically, and ongoing research keeps providing new insights.

Further Reading

Austin, John L. 1962. How to do things with words. London: Oxford University Press.

Bybee, Joan. 1985. Morphology: A study of the relation between meaning and form. Amsterdam: John Benjamins.

Foley, William A., and Robert D. van Valin. 1984. Functional syntax and universal grammar. Cambridge: Cambridge University Press.

Jespersen, Otto. 1924. The philosophy of grammar. London: Allen & Unwin.

Kiefer, Ferenc. 1987. On defining modality. Folia Linguistica XXI(1). 67–94.

Lyons, John. 1977. Semantics, Vol. 2. Cambridge: Cambridge University Press.

Palmer, Frank R. 1979. Modality and the English modals. Cambridge: Cambridge University Press.

Palmer, Frank R. 1986. Mood and modality, 1st edition. Cambridge: Cambridge University Press; 2nd edition,1990.

Portner, Paul. 1999. The semantics of mood. GLOT International 4(1). 3–9.

Sadock, Jerrold M. and Arnold M. Zwicky. 1985. Speech act distinctions in syntax. Language typology and syntactic description, Vol. 1: clause structure, ed. by Timothy Shopen. ed. Cambridge: Cambridge University Press.

Searle, John R. 1983. Intentionality. Cambridge: Cambridge University Press.

von Wright, E.H. 1951. An essay in modal logic. Amsterdam: North-Holland.

Kleanthes K. Grohmann

See also **Aspect; Speech Acts; Tense and Aspect Marking; Tense: Syntax**

Moore and the Gur Languages

The Gur languages (sometimes called Voltaic/ Voltaique languages, especially by French-speaking scholars), a family in the immense Niger-Congo phylum, are spoken in interior West Africa, in Burkina Faso, Mali, Niger, and the northern parts of Côte d'Ivoire, Ghana, Togo, Benin, and Nigeria. The Gur

language family comprises subgroups including Oti-Volta and Grusi. Most of the Gur languages belong to these two subgroups; Oti-Volta in turn includes Gurma and Mabia (Western Oti-Volta). Since the most prominent languages of the Gur family are in the Mabia subgroup, this subgroup is the focus of this brief survey (see Figure 1).

The Mabia languages can be divided into five main groups; the languages within most of these groups are more or less mutually intelligible. Major languages within each group are written in capitals in the diagram. Western Mabia, which includes Dagaare, Waale, Birifor, and Safaliba, is located in northwestern Ghana and adjacent Burkina Faso. Northern Mabia, which includes just Moore and its dialectal forms, is found mostly in Burkina Faso. Central Mabia is located approximately in the middle of the Gur area. There is one overall name for this group, Gurenne, which includes Frafra, Nankani, and Nabit. A subdivision of this group is Mid-Central Mabia. The two languages in this subgroup, Buli and Konni, although geographically close to Gurenne, are linguistically slightly distinct from Gurenne and some of the other Mabia languages; in fact, there is some controversy as to whether this group belongs in Mabia or in Grusi. Here, it is included in Mabia because it has more similarities in vocabulary with Mabia than with Grusi. Southern Mabia, comprising Dagbane,

Mampruli, and Nanuni, is at the south of the Gur area. Southern Mabia too has a subdivision, Mid-Southern Mabia (Hanga-Kamara). Hanga and Kamara, although also found in the south, are quite different from, for instance, Dagbane. The fifth group, Eastern Mabia, is composed of Kusaal and Talni.

This attempt at classifying the Mabia languages does not mean to suggest that there are clear boundaries. Most of the languages, in fact, form a continuum (speakers of adjacent languages can usually understand each other), and variation between the languages is rather gradual. For instance, it is sometimes difficult to say whether Mampruli is more closely related to Dagbane or to Kusaal. A similar situation has been observed for Birifor, Waale, and Dagaare.

Gur languages display some rather interesting typological features. Within typological studies, linguists try to outline the characteristics of sound systems and word and sentence structures in natural languages and then find out how a particular language or group of languages may be classified according to these types of grammatical features.

The Gur languages are marked by a preponderance of consonants and a scarcity of vowels when compared to Indo-European languages like English and French, and they have some consonants with complex articulation that is rare in Indo-European languages.
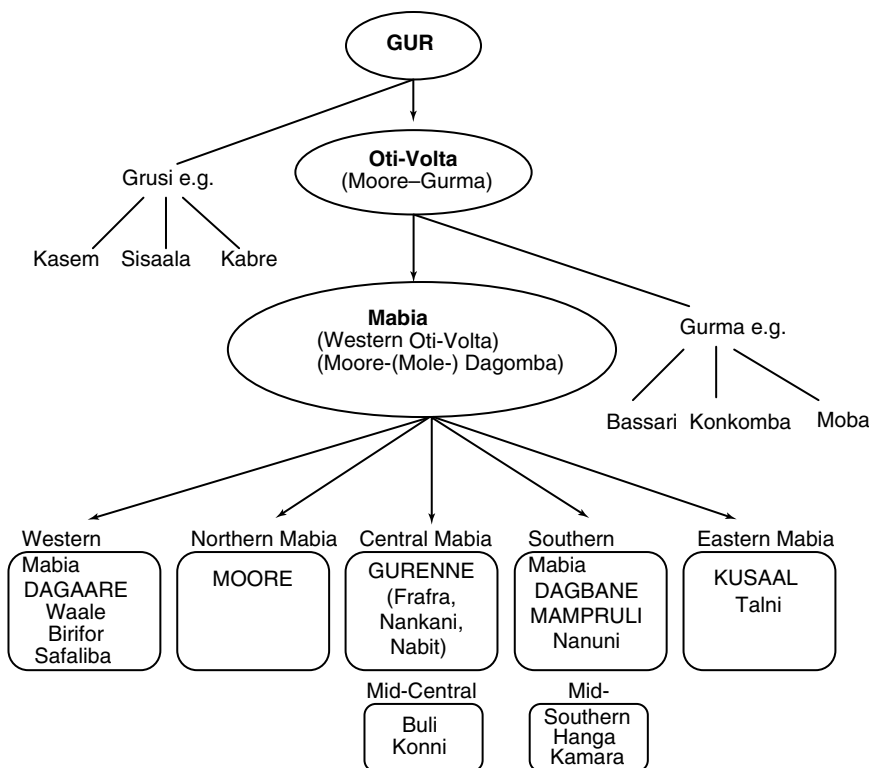
Figure 1

An example would be labio-velar stops, i.e. consonants for which the lips are pressed together as in [b] and the tongue is simultaneously pressed against the back of the roof of the mouth as in [g]. This is actually a common African language feature, even though it is conspicuously absent in languages such as Akan.

With respect to vowels, there is the typological feature of vowel harmony, i.e. the vowels within a word tend to be alike in certain respects and according to specific rules. Not only does this feature distinguish between some of these languages and Indo-European languages it also divides these languages, into harmonizing and nonharmonizing languages. For instance, one difference between Western Mabia languages like Dagaare and Waale and the rest of the group is that these exhibit vowel harmony while the other members of the group do not seem to. It may, however, be possible that vowel harmony existed in the ancestor form of all these languages.

Gur languages are tone languages, i.e. pitch differences within a word can alone change the meaning. Most Gur languages use only two tone levels, i.e. they distinguish between high and low tones. The former are indicated by an acute accent (*á*), while a grave accent (*à*) indicates the latter. These tones serve to express both lexical and grammatical oppositions as in the Dagaare verbs *dá* "push", *dà* "buy".

An important typological feature for these languages is the system of noun classes—nouns belonging to different classes are inflected differently (the masculine and feminine genders of many European languages are examples of noun classes). The manifestation of noun classes is common in Niger-Congo languages but, while these languages usually mark noun classes with prefixes, most Gur languages use suffixes.

Moore

Moore, the language of the Mossi, is the most prominent Gur language in terms of the number of speakers and political importance. It is spoken by about 5 million people in Burkina Faso, where it is the de facto national language, and about 1 million more in Côte d'Ivoire, Ghana, Togo, and elsewhere. Moore shows typical Gur language characteristics. It has two tones, it exhibits vowel harmony, it has a system of noun classes where suffixes indicate the class a noun belongs to, and it has Subject–Verb–Object order. The three sentences below illustrate this structure:

M rîgdá ráágà wá
I going market the
"I am going to the market."

Bòè lá fó mààndà
What that you doing
"What are you doing?"

À yáá tónd tèngá nààbá
(S)he saw our village chief
"He saw the chief of our village."

The language as spoken in Burkina Faso comprises four main dialects with focal points around the towns of Ouagadougou (the capital), Koudougou, Koupela, and Ouahigouya, which are all major provincial centers. Glottochronological studies, which attempt to determine how long ago two or more languages split by comparing the number of very common words they have presumably retained from their ancestral language, indicate centuries of divergence within this language group. They show that Moore, including all its dialects, is more closely related to Dagara, a variant of the Dagaare language, with only two centuries of divergence between them.

Dagbane

Dagbane, including Nanuni, is spoken by about 1.5 million people in Ghana. It is a trade language in and around Tamale, the fourth largest town in Ghana. It is a major language of education and literacy in Dagbon, home of the Dagomba. It is taught in various undergraduate programs at universities in Ghana. Dagbane and Mampruli are mutually intelligible.

Dagaare

Dagaare, including Waale and Birifor, is spoken in northwestern Ghana around towns including Wa, Tuna, Jirapa, Lawra, and Nandom, and in Burkina Faso around towns including Dano and Dissin. Native speakers of Dagaare number up to 2 million. It benefited from early missionary activities and has one of the most extensive arrays of literature in the region. It is taught in primary and secondary schools in Dagao, homeland of the Dagaaba, and in various undergraduate programs at universities in Ghana and Burkina Faso.

Gurenne

Gurenne (Frafra, Nankanne, and Nabit) also has a large number of native speakers, numbering about 1 million. It is the language of Bolgatanga, one of the cosmopolitan towns in northern Ghana. It is fast becoming a lingua franca in northeastern Ghana, where speakers of different languages, such as Kusaal,

Kasem, Mampruli, and Talni, use it in the Bolgatanga metropolitan area.

Kusaal

Kusaal, language of the Kusaasi, is spoken by about 250,000 people in the extreme northeast of Ghana around towns including Bawku. It is an important language of communication in the area. Glottochronological studies show that it is most closely related to Mampruli, with only two centuries of divergence between them.

Mampruli

Mampruli, language of the Mamprussi, is spoken by about 100,000 people around towns including Gambaga, Nalerigu, and Walewale. Mampruli is very closely related to Kusaal and Dagbane, having diverged from them only two centuries ago. Indeed, both languages are quite intelligible to its speakers.

Buli

Buli, the language of the Bulsa, with about 65,000 speakers in and around Sandema, also has some literacy programs. It is closely related to the much smaller language Konni but quite distinct from all the other major Gur languages.

Kasem

Kasem, language of the Kasena, belongs to the Grusi subgroup along with languages including Sisaali, Chakali, Tampulma, Kabre, Vagla, and Mo. It is spoken around towns including Navrongo in Ghana and in adjacent settlements in Burkina Faso. It is spoken by approximately 300,000 people. Kasem, like Dagaare, was one of the first literary languages of northern Ghana, having benefited from early missionary activities.

Sisaali

Sisaali, or Isaaleng, language of the Sisaala, is spoken around the towns of Tumu, Gwellu, and Lambussie in Ghana and in adjoining areas of Burkina Faso. It is spoken by about 200,000 people. Sisaali is scarcely intelligible with the other Gur languages, but many speakers of Sisaali speak Dagaare as a second language.

Konkomba

Konkomba, belonging to the Gurma group along with languages including Moba and Bassari, is also called Likpakpalnli. It is spoken by about 500,000 people. Many of its speakers use Dagbane as a second language in their homeland around Saboba and Zabzugu. The language is taught in primary schools, and there are many literacy programs involving Konkomba.

To sum up, most of the Gur languages are in use for educational purposes in their communities. Some of them serve communities beyond their traditional areas. There are some mass communication programs on radio and TV in these languages in Ghana, Burkina Faso, and other countries. The unrelated Chadic language, Hausa, is often thought to be an effective lingua franca in the Gur-speaking area, but this is hardly a fact. It may have been so some time ago, but currently some of the Gur languages, including Moore, Dagbane, and Dagaare, are fast replacing it as lingua francas in their respective areas of influence. There are serious attempts at functional literacy and mass communication in all these indigenous languages. Awareness is gradually being raised regarding the importance of mother tongues as major languages of mass communication for socioeconomic development in this part of West Africa.

References

Barker, P.1986. Peoples. Languages and religion in Northern Ghana. Ghana Accra: Evangelical Committee.

Bodomo, A.B. 1994. Language, culture, and history in Northern Ghana: an introduction to the Mabia linguistic group. Nordic Journal of African Studies 3(2). 25–43.

Bodomo, A.B. 1997. The structure of Dagaare, Stanford Monographs in African Languages. Stanford, USA: CSLI Publications.

Bodomo, A.B.2000. Dagaare: languages of the world materials No. 165. Munchen, Germany: Lincom Europa.

Goody, J. 1967. The social organisation of the LoWiili. London: IAS.

St. John-Parsons, D. 1958. Legends for Northern Ghana. London: Longmans and Accra: SPC.

Manessy, G. 1975. Les Langues Oti/Volta. Paris: SELAF.

Manessy, G. 1981. Les Langues Voltaiques. Les Langues dans le Monde Ancien et Moderne, ed. by Jean Perrot. Paris: CNRS.

Naden, A.1988. Language, history, and legend in Northern Ghana. Sprache und Geschichte in Afrika 9.

Naden, A. 1989. Gur. Niger-Congo, ed. by Bendor-Samuel. Lanham, MD: Universities Press of America.

Nikiema, E. Moore≅: Languages of the world materials No. 383. Munchen, Germany: Lincom Europa, forthcoming.

Swadesh, et al. 1966 . A preliminary glottochronology of the Gur languages. Journal of West African Languages 3.

Tuurey, G.1987. Introduction to the Mole-speaking community. WA: Catholic Press.

Adams Bodomo

See also **Niger-Congo**

Morpheme

The word *uncountable* is made up of three units: *un-, count,* and *-able*. Each of these units recurs elsewhere in English, for example, in the words *unkind, count-down,* and *acceptable,* respectively. When they recur, they have the same meaning (*un-* means, approximately, 'not' in both *uncountable* and *unkind*) and the same form or a closely related one (in the examples cited so far, the form is identical). Moreover, we cannot subdivide these units into smaller units with these same properties. It is true that we can see two different letters standing for two different sounds in *un-,* /ʌ/ and /n/, but although these have the same form in other words, such as *sun,* they do not have a meaning there: *sun* does not mean 's not.' Lastly, the units *un-, count,* and *-able* between them exhaustively analyze the word *uncountable*: they account for every bit of form; there are no unexplained bits left over.

Let us define units that have all of these properties as 'morphs', although we shall see below that not all scholars require every one of these properties.

When a morph can stand alone as a word, such as *count,* it is called a (potentially) free morph. When, like *un-,* it is found only attached to something else, it is called an (obligatorily) bound morph.

Uncountable is related to the word *uncountability*. Again, we can divide this into morphs as *un·count·abil·ity,* where the *-ity* bit recurs in words like *probability* and *sensitivity,* with a meaning something like 'quality of being'. However, a close comparison with *uncountable* will show that *-able* /əbl/ has changed to *-abil-* /əbɪl/. When the second syllable of *–able* is stressed, as it is whenever *-ity* is added, we find *-abil-* instead of *-able*. The meaning is the same, and the form is still very closely related (it still contains the sounds /ə/, /b/, and /l/), but *-able* and *-abil-* occur in different environments. They are in complementary distribution: where you find one, you cannot find the other. We call morphs that are in complementary distribution in this way 'allomorphs' of the same morpheme. The terminology is entirely parallel to the terminology used in the theoretical descriptions of speech sounds, where one speaks of 'allophones' of the same 'phoneme'. Where there are not several different allomorphs of the same morpheme, as with *count,* whose form is always *count,* we can still say that the single morph *count* instantiates the morpheme {count}, where the braces show that we are treating this as a morpheme.

There are many problems with this notion of a morpheme, and the linguistic community has responded by overtly or covertly modifying the definition of morpheme in several different ways, to such an extent that there are probably few who would adopt the characterization that has been provided here, although they might allow the definition of a morpheme as the minimal unit of grammatical analysis.

Consider the plural markers *in cats, oxen, mice,* and *tempi*. The regular marker on the end of *cats* is not a problem and can be dealt with easily: the word can be analyzed into two morphs, a root *cat* and a suffix *-s,* each of which represents a morpheme. The morpheme for the suffix has three allomorphs, one of which is the /s/ found on the end of *cats*. The first question is: is the same morpheme present in *oxen*? According to the characterization given above, it is not, because there is no form in common between /s/ and /ən/; rather, these are two synonymous morphemes. An alternative view gives more weight to the meaning side of the equation, along with the fact that both are members of the same paradigm of number, and allows /ən/ as an allomorph of {plural} determined by the lexeme (vocabulary unit) *ox*. In *mice,* it is not clear that there are two morphs at all, although some scholars have tried to argue that the root of the word is the discontinuous morph /m—s/ and the plural marker is the diphthong /aɪ/. However, if there is just one morpheme {plural}, then perhaps that morpheme is present in *mice,* even if it cannot be segmented out. The term 'portmanteau morph' is used for a morph that represents more than one morpheme, although some scholars restrict it to instances in which morphemes that are usually represented as independent words are merged, as in French *du,* 'of the (masculine)', which can be seen as a conflation of *de,* 'of ', and *le,* 'the (masculine)'. In the case of *tempi,* there is an alternative plural *tempos,* hence, the *-i* marker and the *-s* marker are not in complementary distribution. However, the fact that they do not contrast is usually accepted as sufficient for them to belong to the same morpheme.

Yet other scholars do not even insist on a constant meaning for the morpheme. Mark Aronoff (1976) argues that because *permit, emit,* and similar words all have irregular nominalizations in *-mission,* the consistent change from *mit* to *mis* is sufficient to establish *-mit* as a morpheme, despite the fact that it cannot be assigned a constant meaning in modern English. For other scholars, this is a matter of etymology (historical development) rather than a

matter of morphology. Joachim Mugdan (1986) provides a survey of definitions of the morpheme.

The problems with the concept of the morpheme are so great that some linguists have given up on the notion entirely. Stephen Anderson (1992) introduces the notion of a-morphous morphology, where the structure of words is determined by the construction of a sound structure on the basis of roots and syntactico-semantic structure, without any recognition of an intermediate level at which there are morphs or morphemes. However, such a stance is still controversial (although it is becoming more generally accepted), and the morpheme is still recognized as one of the fundamental units of linguistic analysis.

References

Anderson, Stephen R. 1992. A-morphous morphology. Cambridge: Cambridge University Press.

Aronoff, Mark. 1976. Word formation in generative grammar. Cambridge, MA: MIT Press.

Coates, Richard. 1999. Word structure. London and New York: Routledge.

Matthews, P.H. 1993. Grammatical theory in the United States from Bloomfield to Chomsky. Cambridge: Cambridge University Press.

Mugdan, Joachim. 1986. Was ist eigentlich ein Morphem? Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 39. 29–43.

LAURIE BAUER

See also **Affixation; Phoneme; Word**

Morphological Typology

Morphological typology is the study of differences among the world's languages relating to the ways in which words are formed from smaller meaningful units referred to as 'morphemes'.

If we examine the formation of plurals in English, then regular plurals like *cats* as the plural of *cat* show the following relation between word and morphemes. *Cats* is a single word, as indicated by the fact that it is a minimal free form, i.e. it is not possible to pronounce *cats* with a pause within the word. In terms of its morphological structure, however, it clearly consists of two meaningful elements, the lexical root *cat* and the plural suffix *-s*. The single word *cats* thus consists of two morphemes and can readily be divided into those two morphemes (*cat + s*). Not all plurals, however, follow this pattern. Thus, the plural of *tooth* is *teeth*, with the plural being marked not by attaching a suffix to the root, but rather by changing the vowel of the root. And occasionally in English, one forms the plural by means of a separate word, as in the nonstandard plural of *who* as *who all*, for instance in *Who all are coming to dinner tonight?*

The different ways of forming plurals in English correspond to the three major morphological types that have been identified by considering many languages around the world. 'Isolating' or 'analytic' morphology refers to a system in which each word consists of only a single morpheme. Vietnamese is a language that comes close to being an ideal isolating language, as can be seen from the following example:

Khi	tôi	đến	nhà	**bạn**	tôi,
when	I	come	house	friend	I

chúng	tôi	**bắt**	đâu	làm	bài.
PLURAL	I	begin		do	lesson

'When I came to my friend's house, we began to do lessons.'

None of these Vietnamese words shows any of the morphological changes that are found in the English translation, such as *come ~ came, begin ~ began,* or *I ~ my;* indeed, even the plural pronoun 'we' is expressed by combining a separate word indicating plurality, *chúng,* with the pronoun *tôi* 'I.' The only exception to strict isolating structure is the verb *bắt đâu* 'begin,' which is a compound word, i.e. a single word made up of two components, literally 'seize' and 'head.' (Note that Vietnamese orthography leaves a space between syllables, and not only between words; the Vietnamese for 'Bolivia' is *Bô li vi a*.)

In contrast to isolating languages, we find 'synthetic' languages, which permit more than one morpheme to combine to form a word. In one type of synthetic language, namely 'agglutinating' languages, the boundaries between the individual morphemes are clear-cut. The classic example of an agglutinating language is Turkish. If we look at the way in which a Turkish noun changes its form to show different numbers (singular vs. plural) and cases (e.g. nominative or dative), we can always easily identify which sequence of sounds is the root, which sequence indicates number, and which sequence indicates case. Thus, the dative plural of *adam* 'man' is *adamlara* 'to the men', which can be broken down into the morphemes *adam + lar + a;* the root is *adam,* and we find this in the nominative singular *adam*

'man' (both singular number and nominative case take no suffix); the plural suffix is *-lar,* and we find this in the nominative plural *adamlar* (*adam* + *lar*) 'men'; the dative suffix is *-a,* and we find this in the dative singular *adama* (*adam* + *a*) 'to the man'. (Note that Turkish has no definite article corresponding to 'the'.) Knowing that the ablative suffix is *-dan,* we can readily form the ablative singular *adamdan* (*adam* + *dan*) 'from the man' and the ablative plural *adamlardan* (*adam* + *lar* + *dan*) 'from the men'. Incidentally, Turkish can in this way string together quite long sequences of suffixes, as in *avrupalılaştırılamıyanlardansınız* 'you are one of those who cannot be Europeanized.' From the noun *Avrupa,* the suffix *-lı* forms the adjective 'European'; from this, the suffix *-laş* forms a verb meaning 'become European'; the suffix *-tır* adds the meaning 'cause to', i.e. 'cause to become European' or 'Europeanize'; *-ıl* makes this passive, i.e. 'be Europeanized'; then *-amı* adds the element of inability, i.e. 'unable to be Europeanized'; *-yan* turns this into a participle, i.e. 'being unable to be Europeanized, one who is unable to be Europeanized'; *-lar* makes this plural and *-dan* adds ablative case, i.e. 'from those who are unable to be Europeanized'; and finally *-sınız* adds the element 'you'.

A second kind of synthetic language is a 'fusional' language, in which the various morphemes fuse together to give a single, unsegmentable whole. We have already encountered an example of fusional morphology in the irregular English plural *teeth,* in its relation to *tooth,* since in this plural the elements root and plural are not segmentable as they are in a regular plural like *cats.* Certain Indo-European languages such as Latin, Greek, and Russian provide good examples of fusional morphology. Whereas in Turkish the markers of case and number are always separable from one another, in Russian the two are inextricably fused. In the nominative, we have the singular *stol* 'table' and the plural *stoly* 'tables'. In the genitive, the singular is *stola* 'of the table' and the plural *stolov* 'of the tables'; it is already clear that the genitive plural suffix *-ov* cannot be segmented into a component that we could identify with nominative plural *-y* and another that we could identify with genitive singular *-a.* The picture is only further compounded as we examine other cases, e.g. dative singular *stolu* 'to the table' vs. plural *stolam* 'to the tables'.

A fourth morphological type is sometimes, but by no means always, recognized, namely 'polysynthetic'. Characteristic of polysynthetic languages is that typically a large number of morphemes are combined in a word, i.e. these languages are, as one might suspect from their name, very synthetic. An example is provided by the Chukchi language spoken on the Chukotka peninsula at the eastern tip of Asiatic Russia, where one finds the single-word sentence *tə-mejŋə-lewtə-pəɣt-*

ərkən 'I have a fierce headache,' whose components are the first-person singular prefix *t-* 'I', the root *mejŋ-* 'big, great', the root *lewt-* 'head', the root *pəɣt-* 'ache', and the imperfective aspect suffix *-rkən* here indicating an ongoing state. (The other instances of ə are inserted to break up consonant clusters.) The possibility of having a single word corresponding to a sentence of several words in most other languages of the world is also often cited as a characteristic of polysynthetic languages. However, it is not really clear whether polysynthetic should be recognized as a distinct type. In the Chukchi example just given, the basic morphological technique is agglutination, and the only unusual feature is the extent to which such agglutination is carried out, combining both lexical morphemes like 'big', 'head,' and 'ache' and grammatical morphemes like first-person singular and imperfective aspect. Polysynthesis is thus arguably just agglutination (sometimes with elements of fusion) carried to an extreme.

Although some languages come close to one of the three ideal types, isolating, agglutinating, and fusional, such as Vietnamese, most languages combine elements of all three types to different degrees. For instance, English is generally an isolating language, and it is quite possible to produce quite long sentences that involve no word consisting of more than one morpheme, e.g. *Every day I see the cat on the wall near the tree.* Agglutinating morphology in English is limited to compounding (e.g. *birthday,* i.e. *birth* + *day*), to the genitive and plural forms of nouns (*cat's, cats*), and to a handful of verb forms (third-person singular present tense *walks,* past tense *walked,* past participle *walked,* present participle *walking*—note that some verbs do have distinct past tense and past participle, e.g. *ate, eaten*). Fusional morphology is limited to a handful of irregular noun plurals (like *tooth ~ teeth*) and a limited number of irregular verbs in the past tense and past participle (e.g. *drink ~ drank ~ drunk*).

The major morphological types can be given a systematic explication in terms of two indices. The 'index of synthesis' indicates how many morphemes are combined into a single word. The 'index of fusion', relevant only if the index of fusion is above 1, indicates how many morphemes are fused together into a single form. By calculating these indices for representative texts in a language, one can assess that language's index of synthesis and index of fusion, thus capturing the intuition that all languages combine the different types to different degrees.

Historically, agglutination can often be shown to result from separate words becoming affixes, as when the Modern English suffix *-hood,* as in *childhood,* derives from an Old English word *hād* meaning 'quality, condition'. Likewise, fusion can often be

shown to have developed from the effect of a lost affix; e.g. a plural suffix *-i*, long since lost, caused the vowel change in nouns like *tooth ~ teeth*.

References

Comrie, Bernard 1989. Language Universals and Linguistic Typology, 2nd edition. Oxford: Blackwell and Chicago: University of Chicago Press.

Greenberg, Joseph H., A quantitative approach to the typology of language. International Journal of American Linguistics 26. 178–94.

Greenberg, Joseph H. 1974. Language Typology: A Historical and Analytic Overview, The Hague: Mouton.

Sapir, Edward. 1921. Language: An Introduction to the Study of Speech, New York: Harcourt, Brace & World.

BERNARD COMRIE

See also **Morpheme; Typology; Word**

Morphology

The central focus of those who study morphology is how language users understand complex words and how they create new ones. Compare the two English words *secure* and *insecurity*. The word *secure* cannot be broken down further into meaningful parts. It is morphologically simple. By contrast, *insecurity* is morphologically complex because it consists of three unanalyzable meaningful components (i.e. in + secure + ity), which linguists call morphemes. The study of the patterning of morphemes within a word and how morphemes combine to form new complex words falls within the domain of morphology. Morphemes are meaningful elements and must be distinguished from units of sound, because a simple morpheme may be complex in its sound structure: the simple morpheme *secure* is complex in terms of sound, consisting of two syllables and six phonemes. The study of the sound structure in language processing is well established. Only relatively recently, however, have psycholinguists begun to examine morphology as a window on how we process words.

The first morpheme in *insecurity* is the prefix (in-), which means approximately 'not'; the second or stem is the adjective *secure*; the third and word final morpheme is the suffix (-ity), which serves to form a noun from the adjective *secure*. The meaning of *insecurity* (viz. 'the state of not being secure') is fully predictable from the meaning of its components. It is semantically transparent. The same prefix occurs in many other words, as does its suffix (e.g. *inactivity*, *impartiality*). Note, however, that not all adjective stems can combine with morphemes such as *in* or *ity*. For example, the stem *abashed* combines with the affixes *un* and *ness* to form the noun *unabashedness* while *continue* combines with *dis* and *ity* to form the adjective *discontinuity*. The fact that nouns like **incontinueness* or **disambiguosity* are impossible

also tells us that there are restrictions on how morphemes combine. At the same time, we can understand novel forms such as *unfaxable*. It is the morphologist's job to discover the general principles that underlie our ability to form and understand some complex forms but not others.

Affixes are morphemes that appear before or after the stem morpheme (viz. prefixes and suffixes, respectively). Affixes may vary quite widely in their productivity, the likelihood that they will appear in new words. Compare the three English suffixes *-ory, -ive,* and *-able*, all of which form adjectives from verbs. The first suffix is almost completely unproductive in Modern English; very few new words with this suffix have been added to the language in centuries. The second occurs in such recent words as *adaptive* and *adoptive*. The third is highly productive: innovations such as *unfaxable* are common. Productivity is usually defined with respect to the extent to which a morpheme is expected to appear in novel forms. For example, if we search a large database for new words (words that do not appear in a large standard dictionary) containing the two morphemes *in-* and *-ity* that we introduced at the beginning of this article, we find very few new words with in- (actual examples include *ineliminable* and *inegalitarian*), but many more with *-ity* (actual examples include *avuncularity* and *deviosity*). We therefore say that *-ity* is more productive than *in-*. However, sometimes researchers define productivity with respect to the total number of words in which a morpheme appears. Again, we find that the affix *able* appears in many more than does the affix *ive*. Analogous to affixes, stem morphemes can differ with respect to their family size, that is, the number of words formed from a particular morpheme stem. For example, many more words are formed from the stem *sist* (i.e. consist, persist, desist, insist, and

approximately 31 derivations) than from the stem *flect* (i.e. inflect, deflect, reflect, and approximately seven derivations).

Depending on whether or not they can stand alone, morphemes are classified into two basic types: free (e.g. *secure*) and bound e.g. *in*, *ity*, *fect*). Affixes are bound because they cannot appear in isolation, but must combine with (be bound to) another morpheme to form a word. Repeated additions of bound morphemes allow for the formation of more complex words as affixes pile one onto another. The word *inconclusiveness* contains three bound morphemes as well as a free stem [conclude] and has been built up in stages from *conclude* by first adding the suffix *-ive* to the verbal stem (with a sound change from d to s), so as to produce *conclusive*, then the prefix *-in* so as to form *conclusive*, and finally the suffix *-ness*, resulting in [[in[[conclude]$_V$ ive]$_A$]$_A$ ness]$_N$. Note that *in* cannot combine with *conclude* but it can combine with *conclusive*. This reflects the strict ordering of stages when complex words are formed. Sometimes, two morphemes can be added to a stem in two different orders, yielding two different meanings. Consider the word *unbalanced*. It can mean either 'not balanced' or 'deranged'. The first meaning results from adding the suffix *-ed to balance* and then adding the prefix *un-,* which, when attached to adjectives means 'not'. The second meaning results from first adding the prefix *un-* to the verb balance, giving us the verb *unbalance*, which means 'derange'. When the suffix *-ed* is added to this complex verb, it creates the adjective meaning, hence 'deranged'.

Finally, among bound morphemes, linguists distinguish inflectional from derivational morphemes. Derivational morphology deals with how distinct words are related to one another; inflectional morphology focuses on the different forms that a word may take, depending on its role in a sentence. English is quite sparse inflectionally, as compared with many other languages (e.g. Classical Greek and most Slavic languages), where each noun, verb, and adjective will have a large number of inflected forms.

The way in which morphemes combine to form complex words also varies across languages. Some languages (e.g. Chinese) have little in the way of combining morphology. Others (e.g. Turkish) are distinctive for the manner in which multiple morphemes occur within a single word. Rules for combining morphemes also vary across languages. In English, morphemes are linked linearly as in words such as *inconclusiveness.* In Hebrew, by contrast, morphemes can be interleaved with one another (e.g. R-G-L combines with -e-e to form ReGeL meaning 'leg' and with m-a-e?- to form mragel? meaning 'spy'). The relation between 'legs' and 'spies' may

seem obscure until one realizes that in biblical times one could only spy by walking around.

Knowledge about words is represented in the mental lexicon. A major research question for psycholinguists is how the mental lexicon represents morphological knowledge. One issue is whether regular forms and irregular are represented differently in the mental lexicon. If regularity is defined with respect to form, we can ask whether words with stems that undergo a change in spelling (and sometimes pronunciation) are represented differently from words whose stem is always regular. That is, are forms such as *run* and *ran* represented differently from forms such as *turn* and *turn*ed? Derivation tends to be semantically somewhat unpredictable: *walker* can mean either 'one who walks' or 'a special support that helps one to walk'. If regularity is defined with respect to meaning, a second issue is whether inflected forms such as *concluded* and derived forms such as *conclusive* are represented in the same manner. Similarly, we can ask whether semantically opaque (or ambiguous) forms (e.g. *walker*) as well as semantically transparent morphological relatives (e.g. *talker*) are represented in a like manner in the lexicon. Within the mental lexicon, some theorists express morphological relatedness in terms of representations that are decomposed and share a constituent morpheme. Other theorists express morphological relatedness in terms of a principle of similarity among full forms.

In the psycholinguistic literature, the classic task for exploring morphological knowledge is the lexical decision task. Letter strings are presented visually. Readers must decide whether each is a real word, and reaction time to decide is measured. Words are usually presented in pairs, a prime and then a target. Typically, prime and target occur in immediate succession. When both prime and target are fully visible, decision latencies to the target are faster when it is preceded by a prime that is morphologically related (e.g. *turned-turn*) than by a prime that is similar in form (e.g. *turnip-turn*) or meaning (e.g. *rotate-turn*) but not morphologically related. Reduced decision latencies following a morphologically related prime as compared to an unrelated prime is termed morphological facilitation and many psycholinguists interpret the effect as evidence that the same base morpheme or stem was activated by the prime and by the target. Typically, morphological facilitation arises both when morphological relatives of the target are regular (e.g. *turned-turn*) and when they are irregular (e.g. *ran-run*), however. Nonetheless, the magnitude of morphological facilitation in lexical decision as well as in other experimental tasks tends to be attenuated when the prime is irregular and involves either reduced form or reduced semantic similarity

with its target. Finally, morphological facilitation has been documented in a variety of languages including those where morphemes are not generally appended linearly (viz., nonconcatenative) such as American Sign, and Hebrew, as well as in many where morphemes are concatenated such as Bulgarian, Dutch, English, French, German, Greek, Serbian, and Spanish. Dimensions of morphological relatedness that can alter the magnitude of morphological facilitation in variants of the lexical decision task include inflection vs. derivation, semantic transparency or opacity (or ambiguousness), and orthographic regularity vs. irregularity.

References

Aronoff, M. 1994. Morphology by itself. Cambridge, MA: MIT Press.

Baayen, H., and A. Renouf. 1996. Chronicling the Times: productive lexical innovations in an English newspaper. Language 72.

Bauer, L. 1983. English word-formation. Cambridge: Cambridge University Press.

Bertram, R., R.H. Baayen, and R. Schreuder. 2000. Effects of family size for complex words. Journal of Memory and Language. 42.

Booij, G., and J. van Marle. Yearbook of morphology. Dordrecht: Kluwer (annual).

Feldman, L. B. (ed.). 1995. Morphological aspects of language processing. Hillsdale, NJ: Lawrence Erlbaum.

Feldman, L. B. 2000. Are morphological effects distinguishable from the effects of shared meaning and shared form? Journal of Experimental Psychology: Learning, Memory and Cognition 26.

Marslen-Wilson, W., L. Tyler, R. Waksler, and L. Older. 1994. Morphology and meaning in the English mental lexicon. Psychological Review 101.

Matthews, P. H. 1991. Morphology, 2nd edition. Cambridge: Cambridge University Press.

Rueckl, J.G., M. Mikolinski, M. Raveh, C.S. Miner, and F. Mars. 1997. Morphological priming, connectionist networks, and masked fragment completion. Journal of Memory and Language 36.

Sandra, D., and M. Taft. (ed.) 1994. Morphological structure, lexical representation and lexical access. Hove, UK: Lawrence Erlbaum.

Spencer, Andrew. 1991. Morphological theory. Oxford: Blackwell.

Zwicky, A., and A. Spencer. (eds.) 1997. Handbook of morphology. Oxford: Blackwell.

LAURIE BETH FELDMAN AND MARK ARONOFF

See also **Affixation; Inflection and Derivation; Lexicon: Overview; Word**

Murrinh-Patha and the Daly Languages

Murrinh-patha is an Australian Aboriginal language spoken on the country's far northern coast, in an area several hundred kilometers to the southwest of Darwin. The language, which now has no significant dialectal variations, is spoken by some 2,000 people, predominantly in the township of Wadeye (formerly Port Keats) and the numerous satellite out-stations and dry season camps that it supplies. With this number of speakers, Murrinh-patha stands out as being one of the few Australian languages not under immediate threat of extinction. Indeed, the Murrinh-patha speech community has been growing rather than diminishing in size over the last few decades, with an expanding population of Murrinh-patha children continuing to learn it as a first language, and with Marringarr-, Mati Ge-, and Marrisyefin- speaking groups from the neighboring regions showing generational shift to Murrinh-patha as they become increasingly integrated into a greater Wadeye community. The language is further supported through a bilingual program in the Wadeye School and language maintenance programs undertaken by the community Language Centre.

Classification

For many years, Murrinh-patha was regarded as something of an Australian isolate, accepted as a member of the Australian language family, but held to belong to no lower level subgroup. In particular, Murrinh-patha was thought to have no close genetic link with any of the dozen or so languages of the Daly River region to its north and east, these being presented by Tryon (1974) as together constituting the 'Daly Family'. The cornerstone to this view of Murrinh-patha's genetic status was the lexical data; Murrinh-patha has at most an 11% shared vocabulary density with any other language against which it has been tested.

Present research is, however, overturning these long-held assumptions. Green (2003) has now made out a compelling case for considering Murrinh-patha as making up a genetic subgroup with the Ngan'gi languages (Ngan'gikurunggurr and Ngen'giwumirri), formerly claimed to constitute a branch of the Daly family. The case is based primarily on formal

correspondences in the core morphological sequences of finite verbs, which Green argues are too matching in their complexities and irregularities to have plausibly come about through anything other than a shared genetic legacy, and which he demonstrates through reconstruction of finite verb paradigms as being systematically derivable from an innovative common parent.

At the same time, Tryon's construction of an overall Daly Family is now also seen as problematic. Green has suggested that in place of the single family proposed by Tryon, the formal evidence establishes five separate Australian subgroups in the region. These are given in Table 1. (Identifiable branches of each subgroup are listed on separate lines.) Green claims that the five subgroups cannot convincingly be related together as a single genetic unit, and argues that the similarities that Tryon considered to be diagnostic of membership in the Daly Family are better accounted for either diffusionally or as genetically inherited features shared with a wide range of northern Australian languages.

TABLE 1 Genetic Subgroups in the Daly River Region

Subgroup	Principal Language Varieties
Anson Bay	Batjamalh (aka Wadjiginy), Kendjerramalh (aka PunguPungu)
Northern Daly	MalakMalak, Tyeraty, Kuwema
Eastern Daly	Matngele, Kamu
Western Daly	Marrithiyel, Marrisyefin (aka Marri Tjebin), Marri Ammu, Marringarr, Mati Ge (aka Magati Ge) Marramaninydyi, Marranunggu (aka Merranunggu and Warrgat), Emmi-Menhdhe
Southern Daly	Murrinh-patha, Ngan'gikurunggurr, Ngen'giwumirri

Areal Features

Like the majority of languages spoken in Australia's central far north, Murrinh-patha and the Daly languages are of the 'polysynthetic' structural type, and are categorized within Australianist typology as belonging in the 'non-Pama-Nyungan' and 'prefixing' groups. They possess complex verbal structures built up through the addition of strings of prefixes and suffixes to the lexical roots. Many of the affixes are suppletive in form and portmanteau in nature, simultaneously encoding a number of grammatical categories. Equally, it is not uncommon for grammatical categories, such as person, number, and tense, to be marked discontinuously via different affixes at different points in the verb, as illustrated in example 1.

Example 1. Murrinh-patha (Walsh 1993a,b)
ma-nanku-ma-purl-nu-ngintha-nga-ni
1sgS.Hands.Fut-2dlO-hand-wash-Fut.   Positive-2dl.non-male.mixed sub-section-1sgS.Fut-Sit
I will keep on washing the hands of you two (who are not siblings and one or both of whom are female).

Pronominal Indexing

Verbs obligatorily index core participants such as 'subject' (I saw you) and 'object' (I saw you) by bound pronominal prefixes. Most languages additionally allow for the verbal cross-referencing of other kinds of participants, such as 'goals' (I said it to her), 'benefactives' (I did it for her), and 'adversatives' (My wife ran away on me). Pronominal indexing is shown in example 2, from Ngen'giwumirri (< Southern Daly), where subject, object, and adversative are indexed on the one verb. Observe that the subject pronoun also encodes a tense category, in this case the perfective. Variation of the subject pronoun for tense is found in all languages, typically as the historical result of fusion with a formerly discrete tense affix.

Example 2. Ngen'giwumirri
Danginy-nyi-fime-ngidde-wurru
3sgS.Poke.Perf-2sgO-give-1sgAdv-bad
She gave it away to you against my wishes.

Verbal Classification

Many languages of Australia's central far north share the further characteristic of forming their verbs with not one, but rather two, root-like elements. This two-part structure typically involves the pairing of a relatively inert root (or 'co-verb'), the element that provides the main lexical information for the verb, with a root that hosts the core grammatical affixes. The lexical roots form an open class, while the 'grammatical' roots belong to a closed class, normally numbering less than 50. The two-part verbal structure is thought to be an ancient diffusional feature. And while in some other parts of Australia's north the grammatical root has synchronically no clear semantic value, in the Southern and Western Daly subgroups, it functions as a classifier of the verbal action. The Southern Daly languages all have more than 30 of these classifiers, while the Western Daly varieties have around 20 or so. Verbal classification may simply involve specifying the relative orientation of the subject, as with the intransitive classifiers in examples 3 and 4.

Example 3. Marrithiyel
ga-wu-fifi-nysyen-a
3sgS.Realis-stand-smoke-then-Past
He was standing up smoking then.

Example 4. Marrithiyel
ga-fulh-fifi-nysjen-a
3sgS.Realis-lie-smoke-then-Past
He was lying down smoking then.

Verbal classification may also be relatively abstract in nature, as illustrated by the transitive classifiers in examples 5 and 6 below. In example 5, the grammatical verb root functions to conceptualize the action as performed at the end of an elongated instrument. Contrast this with example 6, in which the replacement of the grammatical verb-root achieves a different schematic conceptualization, this time of the action as performed along the shaft, rather than the end, of an elongated instrument.

Example 5. Marrithiyel
sjendi-gin gu-mun-git-a
spear-Instrumental 3sgS.Realis-jab-sever-Past
He cut it with a spear (that is, by jabbing at it with the end of the spear).

Example 6. Marrithiyel
sjendi-gin gi-ny-git-a
spear-Instrumental 3sgS.Realis-swing-sever-Past
He cut it with a spear (that is, by swinging at it with the shaft of the spear).

These rich classificatory systems are not, however, found right across the Daly region. The Northern and Eastern Daly languages, for example, have much smaller and simpler systems, with just a single productive transitive classifier combining with a few intransitives, while the Anson Bay subgroup shows only recent and rudimentary developments in this direction.

When compared to the other languages of the Daly, and to northern Australia generally, the classifying verb structures of the Southern and Western Daly languages reveal two aberrant features. Firstly, they exhibit a tight morpho-phonological binding between co-verb and the inflected grammatical root. Secondly, they show an innovative ordering, placing the co-verb after the inflected grammatical root rather than preposed to it. Reid (2003) argues that these shared features result from recent diffusion rather than a shared genetic legacy, demonstrating how the Southern Daly languages Ngan'gikurunggurr and Ngen'giwumirri have acquired them only within the last hundred years.

Free pronouns

The Daly languages have freeform pronoun systems that are interesting for several reasons. Firstly, they tend to be complex by virtue of grammaticizing multiple nonsingular number categories. Some languages have singular/dual/trial/plural systems, while others have singular/dual/paucal/plural systems. Secondly, some have developed systems of pronouns where the categories of number and person have become integrated with other categories of information. For example, in Murrinh-patha features of the subsection system, and a gender categorization that hinges on the minimal inclusion of at least one female within a group determine pronominal choice. As can be seen from Table 2 (from Walsh 1976), this type of complexity results in nine Murrinh-patha pronouns corresponding to English 'we'.

Nominal Classification

All the Daly languages have at least a few generic nouns, such as 'meat', 'vegetable food', and 'fire', which are regularly placed in front of specific nouns to encode salient cultural categories. In Murrinh-patha, this has become extended into a system in which the category membership of all entities is obligatorily encoded by one of around a dozen NP initial generic nouns.

In an adjacent bloc of Daly languages, encompassing Ngan'gikurunggurr, Ngen'giwumirri, and all but the northernmost Western Daly varieties, the systems have the same number of categories, but have become more grammaticized, displaying agreement phenomena and reduction of the independent generic. Sometimes, agreement is marked by bound forms attached to nouns as well as modifiers such as adjectives or demonstratives. In other cases, noun class assignment is marked by freeform generics that precede specific nouns and also precede the modifiers. Both these possibilities are demonstrated in the Marrithiyel examples below.

Example 9. Marrithiyel
a-madi a-ngelfu
animal-barramundi animal-many
many barramundi fish

Example 10. Marrithiyel
Wadi meri wadi gutlik wadi gan
Male man male blind man this
This blind man

Noun class phenomena in Daly languages have proved theoretically interesting by providing a perspective on the historical development of bound class from freeform to proclitic to prefix. They have also contributed to theorizing about the process by which agreement phenomena develop (Reid 1997) and to considerations of the nature of the distinction between noun class and noun-classifying systems (Green 1997).

TABLE 2   Murrinh-Patha Freeform Pronouns

	1st Inclusive	1st Exclusive	2nd Person	3rd Person	
Singular		ngayi	nyinyi	nukunu	Male
				nigunu	Female
Dual	neki	ngankuninda	nankuninda	peninda	All male, mixed subsections
		ngankunginda	nankunginda	peninginda	At least one female, mixed subsections
		nganku	nanku	piguna	All same sub section
Paucal	nekineme	ngankuneme	nankuneme	peneme	All male, mixed subsections
	nekingime	ngankungime	nankungime	peningime	At least one female, mixed sub sections
Plural	nganki		nanki	pigunu	All same subsection

TABLE 3   Ngen'giwumirri Phonemic Inventory

	Bilabial	Alveolar	Palatal	Velar
Voiceless stop	p	t	ty	k
Voiced stop	b	d		
Fricative	f	s	sy	g
Nasal	m	n	ny	ng
Lateral		l		
Rhotics		r (approximant) rr (trill)		
Glide	w		y	

Phonology

Australian languages generally lack phonemic fricatives, and have just the single series of stop. The Daly region shows a significant departure from this pattern. With the exception of Anson Bay, all the subgroups show at least some phonemic voicing contrast. In addition, in Ngan'gikurunggurr, Ngen'giwumirri, and the Western Daly varieties, there are phonemic fricatives, with a resultant three-way voiced stop, voiceless stop, and fricative contrast in the bilabials, alveolars, and (in some languages) palatals. The phonemes of Ngen'giwumirri, showing the atypical obstruent set in an otherwise standard Australian inventory, are given in Table 3 in their practical orthography.

References

Evans, N. (ed.) 2003. The Non-Pama-Nyungan languages of Northern Australia: comparative studies in the continent's most linguistically complex region. Studies in Language Change 552. Pacific Linguistics: Canberra.

Green, I. 1989. Marrithiyel. Unpublished Ph.D. thesis. Australian National University: Canberra.

———. 1997. Nominal classification in Marrithiyel. In Harvey and Reid.

———. 2003. The genetic status of Murrinh-patha. In Evans.

Harvey, M. and N. Reid (eds.) 1997. Nominal classification in Aboriginal Australia. Amsterdam: John Benjamins.

Reid, N.J. 1989. Ngan'gityemerri. Unpublished Ph.D. thesis. Australian National University: Canberra.

———. 2000. Complex verb collacations in Ngan'gityemerri: a non-derivational mechanism for manipulating valency alternations. Changing valency, ed. by R.M.W. Dixon and A. Aikenvald. Cambridge: Cambridge University Press.

———. 2003. Phrasal verb to synthetic verb: recorded structural change in Ngan'gityemerri. In Evans.

———. 1997. Class and classifier in Ngan'gityemerri. In Harvey and Reid.

Tryon, D. 1974. The Daly family languages, Australia. Series C, No. 32. Canberra: Pacific Linguistics.

Walsh, M.J. 1976. Murinypata of North-West Australia. Unpublished Ph.D. thesis. Australian National University: Canberra.

———. 1993. Languages and their status in Aboriginal Australia. In Walsh and Yallop.

———. 1993. Classifying the world in an Aboriginal language. In Walsh and Yallop.

———. 1996. Vouns and nerbs: a category squish in Murrinh-patha (northern Australia). Studies in Kimberley languages in honour of Howard Coate, ed. by W. McGregor, pp. 227–52. München: Lincom Europa.

Walsh, M., and C. Yallop (eds.) 1993. Language and Aboriginal Australia. Canberra: Aboriginal Studies Press.

NICHOLAS REID AND IAN GREEN

See also **Australia**

N

Nahuatl and Uto-Aztecan Languages

The Uto-Aztecan (also spelled Uto-Aztekan) language family is comprised of eight branches, spread across the Great Basin, into the northwest, California, the southwest of the United State, and through Mesoamerica. The major groups are Numic, Tübatulabal, Hopi, Takic, Pimic, Taracahitic, Corachol, and Aztecan.

Pre-European contact Numic peoples ranged through the western portion of the United States into Alberta and Saskatchewan; postcontact, their range contracted. Today, groups live in Wyoming, Idaho, Oregon, Utah, California, Colorado, Nevada, Arizona, and New Mexico. Numic languages include Comanche and Shoshoni (also written Shoshone), Kawaiisu, Ute, Northern and Southern Paiute, and Mono. Although covering a wide area, languages of this group remain quite similar in lexicon and morphology, suggesting a relatively recent dispersal. Miller (1983) follows Lamb (1958) and Fowler (1972) in surmising that the initial communities inhabited the southwestern part of the Great Basin, which began to separate into northern and southern groups about 2,000 years ago, moving out to their greatest extension around 1,000 years thereafter. Today, Paiute and Shoshoni remain vigorous languages. The Yerington band dialect of Paiute is expanding its forms and domains of usage. Most other Numic-speaking communities are shifting to English as a first or dominant language; this shift is complete for Mono and Kawaiisu.

The Tübatulabal branch consisted of a single family. It became extinct at the end of the twentieth century.

Hopi remains a viable language, spoken in northeast Arizona, Utah, and New Mexico. Eighty percent of ethnically ascribed Hopi speak their language. The Hopilavayi Project of the Hopi Cultural Preservation Office has designed materials to help communities teach the language within the homes, villages, and schools.

Benjamin Whorf relied heavily on his study of Hopi in developing his theory of linguistic determinism. Ekkehart Malotki, working closely with Michael Lomatuway'ma, has published seven collections of Hopi tales, presented in bilingual format. He has also explored the encoding of time in the Hopi language, the basis for much of Whorf's speculation.

The Takic branch, situated in Southern California, consists of Cahuilla, Cupeño, Luiseño, and Serrano. Each of these languages has fewer than 50 speakers; only one speaker was reported for Serrano in 1994. Although California has had a mentor-language-learning system, which pairs young Native Americans with fluent speakers of their languages for total immersion, some elders prefer not teaching the language to teaching one that they must fill with neologisms.

The Pimic branch (also called Tepiman) includes Pima, Papago, and Tepehuan. Papago-Pima is also known as O'odham, O'othham, Nevome, Nebome, and Upper Piman. The term O'odham means 'We, the People'. The Pimas of the Gila River area call themselves *ʔakimel ʔoʔodham* 'river people', while the Papago to the south are *Tohono ʔoʔodham* 'desert people'. Through the early 1900s, Pima elders continued to record their life histories on calendar sticks. The writing system appears to have been a mixture of personal symbols and ideographs. Most of these records were destroyed upon the authors' deaths. The languages are now recorded in the roman alphabet. The

Tohono O'odham have K-12 bilingual education. Two Piman languages, both referred to as Pima Bajo, are spoken in Mexico, Chihuahua, and Sonora. Speakers are just becoming bilingual in Spanish. Tepehuan, divided into three major dialects, is spoken in Durango, Chihuahua, and, until recently, in northwestern Jalisco, and has approximately 35,000 speakers. Tubar, once spoken in Chihuahua, is now extinct.

Taracahitic consists of Guarijio, c. 5,000 speakers in the western Sierra Madre; Tarahumara, in Chihuahua, with around 72,000 speakers; and the Cahitic languages: Mayo, in southern Sonora and northern Sinaloa; Ópata in the Districto Federal and the state of México; and Yaqui, with speakers in Sonora and in Arizona. While Ópata is moribund, Mayo and Yaqui continue to be highly used, and are similar in structure and lexicon. Yaqui is tonal, while Mayo is not.

Corachol is made up of Cora, spoken in Nayarit and in a pocket of western Colorado, and Huichol in northwestern Jalisco. All the languages of this group are tonal.

The Nahuatl branch, also called Aztec, Mexicano, and Nahua, is robust, with around 1.9 million speakers. Precontact Nahuatl-speaking communities were scattered from the southern Sonoran desert through Salvador and Honduras. The Spanish crown made Nahuatl an official language for legal proceedings. Many non-Nahuatl communities retained Nahuatl scribes to file their land claims, wills, and petitions. Spanish priests wrote excellent dictionaries and grammars of the language; a Nahuatl speaker trained by the Jesuit priest/linguist Horacio Carochi composed a series of 'dialogues' meant to serve the priests as a model for polite discourse (Karttunen and Lockhart 1987). In addition, many Nahuatl town scribes recorded local histories in annals; although postcontact and written in the roman alphabet, many of these annals record earlier events. That of Quauauhtitlan covers 635–1519 CE; that of the Chichimecs spans 1116–1544 CE. Some or all of these documents may have been based on earlier nonalphabetically recorded texts. Precontact Nahuatl writings are predominatly picto- and logographic. There is evidence that some signs in late texts were being used for their phonetic value, rather than as illustrations of the words or concepts they represented, perhaps indicating a shift toward a more syllabic or phonetic writing system. Surviving precontact writings are dynastic histories, annals, maps, divinatory guides, or combinations of these.

Nahuatl schools for noble youths were often taken over by priests, postcontact. Early texts produced by young men trained in these schools occasionally retained logograph-like illustrations or marginalia. Fray Bernardino de Sahagún, working with a cadre of Nahuatl scholars, produced a compendium of cultural exegeses, which he entitled *General history of the Things of New Spain*. Treatises cover precontact religious practices, sacred songs, origin tales, divination, philosophy, astronomy, political structure, occupations, plants and animals, and the Spanish invasion. Nahuatl intellectuals in the early colonial period, among them Chimalpahin and Tezozomoc, wrote political and historical essays, and composed poetry and songs. Major tropes include parallelism (morphological, syntactic, and semantic), metaphor, couplets-for-class (naming a class by two prime examplars, as *cuāuhtli, ōcēlōtl* 'eagle, jaguar', the two major warrior societies standing for 'warriors'; *tōchtli, mazātl* 'rabbit, deer', two prime wild animals, standing for the 'wild'), and inversion. Parallelism is a common feature in indigenous literatures of the Americas, as are the couplets-for-class. Inversion is more particular to colonial Nahuatl canons. In inversion, a word or phrase is used to refer to its 'inverse' or antonym. Thus, *cuāuhtli, ōcēlōtl* 'eagle, jaguar' may be used to refer to a commoner, a farmer, rather than to the elite warriors; *pilli* 'child' is also the term for a person of the nobility.

Legal and ecclesiastical records were still kept in Nahuatl through the eighteenth century. In the mid-twentieth century, Mexican intellectuals began a Nahuatl revitalization, publishing poetry and prose again in Nahuatl. UNAM publishes a journal *Tlalocan*, primarily dedicated to Nahuatl studies. The Mexican Ministry of Education sponsors bilingual education in Nahuatl communities. A full range of teaching materials is available for K-12. Mexican-Americans have also taken up Nahuatl as a symbol of identity. Nahuatl words, written and spoken, are used emblematically by populations in California and, to a lesser extent, Arizona.

Pipil is the Nahuatl variety spoken in El Salvador. Official estimates have long placed this language close to extinction. However, a 1979–1980 survey by the Universidad Centroamericana José Simeón Cañas revealed that there were still viable populations of speakers in the three northernmost departments. With the resolution of the civil war, a bilingual education project was established. Materials and teachers from sister communities in central Mexico helped jumpstart the schools in Pipil-speaking communities.

Proto-Uto-Aztecan had a relatively simple consonant inventory: p, t, č, k, kw, ʔ, s, h, m, n, w, y, and five vowels. Classical Nahuatl had long and short i, e, a, and a back vowel [o~u]. Some modern languages within the family are tonal, but tone appears to have been developed from glottal stops and consonantal sandhi. Most languages of the group have fixed stress.

Morphologically, languages of this group tend to share the use of an absolute suffix, although they

vary with respect to its ubiquity. Some languages have only vestiges of this affix in citation forms. In others, such as modern spoken Nahuatl, the affix appears when noun stems are neither possessed nor initial elements of compounds. Compounding is a productive mechanism for word formation, as are cliticization and affixation. In Cora, the verb word may be quite long, as habitual objects, goals, and instruments may be incorporated into the verb word. Verb–verb, verb–adverb compounds are also common. Huichol has productive compounds; Mayo has relatively shorter words, and Classical Nahuatl combined compounding, with elaborate affixation. A verb word in Classical Nahuatl could contain a negator, an incorporated adverbial, an antecessive marker, additional time adverbs, subject pronouns, object pronouns, movement markers, reflexives, incorporated noun objects/goals, reduplication (for intensity or distribution), incorporated indefinite objectmaker, further reduplication, a verb root, a series of derivations affecting transitivity, a connective, one or more auxilliary verbs, aspect marker, and a number marker. Nouns, although less complex, could be elaborate as well, containing reduplication, one or two roots, derivational affixes, and either an initial possessive or a final absolutive maker. Common postpositions marked both positive and negative affect, location, abundance, position, or manner.

Modern Nahuatl retains most of these possibilities, although the extended alternation of transitivizing and intransitivizing verbal suffixes found in courtly colonial Nahuatl writings does not appear in modern honorific usage. The compounding potential of Nahuatl has made the formation of neologisms relatively easy. This potential is open to the other languages of this family.

Classifiers for nouns or for counting are relatively undeveloped within Uto-Aztecan. Most languages have a few. Classical Nahuatl had seven. Modern spoken Nahuatl varieties seldom use more than three. This may be related to the use of Spanish number expressions for most counting.

Proto-Uto-Aztecan appears to have placed verbs following the complement or direct object (SOV). Many modern languages have experienced some reordering in main clauses, while retaining the early order in subordinate clauses. Most languages in this family have postpositions and, at least, allow Subject–Object–Verb order today. Preposing an argument or clause is used for emphasis. Relative clauses may occur without a relative pronoun. Subordination is often by juxtaposition. In Proto-Uto-Aztecan, subject noun phrases of subordinate clauses were inflected as objects. Most languages of the family retain this trait. Nouns and adjectives may serve as predicates, with a 'zero' copula. Generic action may be shown

with indefinite actor pronoun affixes on the verbs. Number can be marked in several ways in Uto-Aztecan languages, including use of numerals, inflection on nouns and pronouns, on verb forms, and on pronominal verb affixes. Languages within the family differ with respect to the obligatoriness of number marking and agreement. Classical Nahuatl did not require number agreement between pronominal affixes and coreferent noun phrases, although increasing animacy brought increasing agreement. Modern spoken varieties of Nahuatl more consistently mark plural on nouns, both animate and inanimate. There is also more agreement between pronominal verb affixes and their coreferent noun phrases. Classical Nahuatl had a variety of mechanisms for inflecting for plural, including reduplication. Modern Nahuatl retains examples of earlier plural classes in fossilized forms. The productive morphemes are now restricted to a few forms for general pluralization, diminuitives, and vocatives. Hopi may mark the plurality on nouns by reduplication. Nouns and verbs in Hopi may also carry plural suffixes. Some Hopi verbs have suppletive stems for the plural. Hopi also makes a distinction between singular, dual, and plural for subjects. Some speakers retain a separate dual noun suffix; others do not, but the singular, dual, plural distinction is shown by agreement on the verb. A singular noun phrase with singular verb agreement is interpreted as singular. A plural noun phrase with a plural verb is plural, while a plural noun phrase with singular agreement is dual.

The Uto-Aztecan family is large and diverse; member languages are spoken from the Canadian border to El Salvador. The family retains many typological and structural features, with enough shared lexicon to allow fairly strong reconstructions. Suárez (1983: 28) gives a glottochronological estimate of 5,100 years of separation.

References

Anderson, Arthur J.O., and Charles E. Dibble. 1982. Florentine codex: general history of the things of New Spain, Fray Bernardino de Sahagún. Santa Fe, New Mexico: The School of American Research and the University of Utah.

Campbell, R. Joe, and Frances Karttunen. 1989. Foundation course in nahuatl Grammar, Vols. 1 and 2. Missoula, MT: The University of Montana.

Fowler, Catherine S. 1972. Some ecological clues to protonumic homelands. Great Basin cultural ecology: a symposium, ed. by Don D. Fowler. Desert Research Institute Publications in the Social Sciences 8. 105–21.

Jaquith, James R. 1970. The present status of the Uto-Aztekan languages of Mexico; an index of data bearing on their survival, geographical location, and internal relationships. Greeley: Museum of Anthropology, University of Northern Colorado.

Kalectaca, Milo. 1978. Lessons in Hopi, ed. by Ronald W. Langacker. Tucson, AR: The University of Arizona Press.

Karttunen, Frances E. 1983. An analytical dictionary of Nahuatl. Austin: University of Texas Press.

Karttunen, Frances E., and James Lockhart. 1987. The art of Nahuatl speech: the Bancroft dialogues. Los Angeles: UCLA Latin American Center Publications, University of California.

Lamb, Sydney. 1958. Linguistic prehistory of the Great Basin. International Journal of American Linguistics 24(2). 95–100.

Mason, John Alden. 1950. The language of the Papago of Arizona. Philadelphia: University Museum, University of Pennsylvania, 1950.

Miller, Wick. 1983. Uto-Aztecan languages. The handbook of North American Indians, 113–24. Washington: Smithsonian Institution.

Press, Margaret Lauritsen. 1979. Chemehuevi: a grammar and lexicon. Berkeley: University of California Press.

Suárez, Jorge. 1983. The Mesoamerican Indian languages. Cambridge: Cambridge University Press.

Judith M. Maxwell

See also **Mexico; United States; Whorf, Benjamin Lee**

Naming

Names have always been seen as a very special class of linguistic signs, and naming as a very special linguistic process. Many cultures and societies associate names and naming with magic and taboo. A famous example is the tale of Rumpelstiltskin:

> The goblin Rumpelstiltskin helps a poor maid to become queen but demands her first child in return. When she gives birth and refuses to give up her child, she can only break the goblin's power over her by discovering his secret name. When she finally does, Rumpelstiltskin explodes with anger and rage and his power is broken forever.

The relationship between person and name is so close that in many cultures saying the name of a person is believed to result in serious harm. Some Australian Aboriginal groups bestow two names on each person: one that is commonly used, and one that is only known to the fully initiated and is only uttered on the most solemn occasions and even then only in a whisper (for further examples of magic and taboo associated with names, see Frazer 1993). This essay provides an overview of onomastics (as the academic discipline of name study is called) by reviewing the status of names as linguistic signs, describing the uses of naming, and discussing three categories of names in more detail, namely personal names, place names, and brand names.

At first glance, the difference between names or 'proper nouns' and common nouns seems straightforward enough: the former have unique reference and denote individuals, while the latter have generic (or 'common') reference and denote classes of entities. This means that all referents of a common noun such as *house* have something in common (walls, windows, roofs, etc.), and a sentence such as *Houses are important for humans to live in* makes sense. However, this is not true for proper nouns, and therefore they normally do not allow for plurals. It is true that names such as *John* or *Newcastle* may have many different referents (John Smith, John Baker, John Howard, etc.; Newcastle, UK; Newcastle, Australia; etc.). But even if a name refers to more than one single extralinguistic referent, those referents do not have anything in common. Therefore, sentences such as *Johns make good lovers* or *Newcastles are good places to live in* do not make sense. This seemingly clear-cut difference between proper nouns and common nouns on the basis of unique and common reference can be attacked on a number of grounds, however. First, on closer inspection the various referents of *John* and *Newcastle* do have some things in common: Johns will be male; will have been born into an English-speaking environment; and, given that most personal names are subject to fashion, one can even make an educated guess about the age of many Johns. Second, there are a number of common nouns whose reference is also unique: *sun, moon, hell,* etc., all refer to a unique extralinguistic referent, although they are not usually considered names.

Other criteria that have been suggested for distinguishing between proper nouns and common nouns are related to sentence structure, translation, and context. The syntactic criterion that defines names (in English) is the absence of a determiner: while common nouns take determiners (*a/the house*), proper nouns do not (*a/the John, a/the Newcastle* are ungrammatical expressions). However, not all common nouns can take the full range of determiners: *a music,* for example, is not a possible expression in English. At the same time, some proper nouns require a determiner (e.g. *The Hague, the Buddha, the Potomac*). With brand names, the distinction is lost completely, as the determiner contrast can be used here to distinguish between the name of a company and the name of a product, as in *Onassis bought Ford* vs. *Onassis bought a Ford.*

With respect to meaning and sentence structure, proper and common nouns clearly form a gradient, with prototypical cases at either end and many fuzzy ones in between. Some scholars have argued that translatability (of common nouns) and nontranslatability (of proper nouns) can provide the only clear distinction. However, counterexamples are easy to find, even if they are rather rare: e.g. *New York* can be translated into German as *Neu York* and into Spanish as *Nueva Yorca*, or German *Schwarzwald* becomes *Black Forest* in English. Ultimately, the intention and perception of language users is the only way to distinguish between common nouns and proper nouns. However, this distinction is no longer systematic and inherent in the language, but determined by context: whether any given noun is considered a name or not would thus shift from context to context. The shifting use of words as proper and common nouns is best exemplified with brand names: from the producers' perspective, brand names have to be unique and distinctive. At the same time, they refer to classes of items that are more identical than any two *houses* are ever likely to be. Indeed, with the widespread practice of *badge-engineering* (diversifying your product range by selling the same item under different names), the name may even be the only difference that exists across different classes of items. With respect to context, there are two important uses to which names can be put. First, they can be used to identify and individualize their referents. It might be argued that this purpose could be more rationally achieved through numerical coding, but names are more memorable and much richer in connotation. Second, personal names are used to address people. Together with terms of address, names allow for rich combinatorial possibilities of expressing the social relationship between speaker and addressee (e.g. Carrie, Carol, Carol Walker, Ms. Carol Walker, Ms. Walker, Dr. Walker, Aunt Carol). Only personal names are normally used in forms of address, and these will now be described in some more detail, followed by place names and brand names.

Personal names are names used for people, but names for pets are also sometimes included in this category. In many parts of the world, personal names consist of a 'given' part and a 'family' part, which tend to be fundamentally different in that given names are usually a matter of choice while the surname runs in the family. The surname is largely conventionalized in that it is usually the father's family name in many parts of the world (although recent societal changes have led to an increase in the mother's name or a combination of parental family names also being used). Research questions concerning these two categories of personal names differ significantly. Questions about first names include the following: (1) How are they formed? Female first names, for instance, often follow a pattern of ending in

-a, as in *Lydia*, *Moira,* or *Sonya*. (2) Where do first names come from? In English-speaking societies, a significant source of first names has traditionally been the Bible (e.g. *Mark, Michael, Peter*). (3) What patterns of sex-specific names are there? For instance, the pool of boys' names has tended to be much more stable over the years than the pool of girls' names. In recent years, there is also an increasing trend toward gender-neutral names such as *Ashley*, *Taylor*, or *Madison*. (4) How has first name choice developed historically and what naming fashions are there? Less than a century ago, the pool of first names was much more limited than it is today, now that parents wish to express their child's individuality as uniquely as possible. (5) How and why do people choose particular names? Reasons include family tradition, 'hero worship' (e.g. naming a child after a celebrity), expressing a wish for the child's future or its personality (e.g. *Victoria, Rose*), or sound appeal. (6) Is there a link between name and personality? Such links undoubtedly exist but it is almost impossible to say whether people live up (or down) to their names or whether they are perceived in particular ways because of the names they carry.

Interest in family names, on the other hand, is often genealogical, and records of family names are used to trace the settlement history of an area or the migratory history of people. Other concerns include the etymology of family names, the pattern of their transmission, and laws regarding naming. Further areas of inquiry into personal names are the use of pseudonyms, naming in literature, name change, nicknames, and their use in functions that are not directly related to naming. These include generic forms such as *Jane Doe* or *John Doe*, metaphoric uses (e.g. *John Bull, Uncle Sam*), use in idioms and proverbs (e.g. *jack-in-the-box, all work and no play makes Jack a dull boy*), and metonymic uses, which are most often found in various specialized registers (e.g. *Mercedes, Creutzfeldt-Jakob disease*).

Place names include names for settlements, rural areas, rivers, and streets. The study of place names is relevant to the work of dialectologists, archeologists, and historians because place names can provide valuable clues to the settlement history of an area, in addition to or in the absence of other linguistic and material records. Britain's Celtic (e.g. *Thames*), Anglo-Saxon (e.g. *Winston*), Latin (e.g. *Chester*), Viking (e.g. *Derby*), and French (e.g. *Bellevue*) names reflect the successive waves of conquest that have shaped the history of the island. Consequently, the main research interest is in the etymologies of place names. However, the formation of place names and their meanings are also of interest. Place names typically refer to the form of a place (e.g. *Hillsdale*), its geological quality (e.g. *Rocky Mountains*), its situation in relation to water (e.g. *Tennant's Creek*), its position (e.g. *West Ryde*), its flora

(e.g. *French Forest*), or its fauna (e.g. *Emu Plains*). Furthermore, place names often refer to human uses to which the land was put (e.g. *Hyde Park*), they commemorate events (e.g. *Breakfast Creek*), or they are named after deities or people (e.g. *San Francisco*).

Brand names have often been excluded from onomastic research on the grounds that they are closer to common nouns than to personal and place names. However, brand names are fascinating, both in form and in meaning, because they are more consciously created than any other words. In addition, there often exist good records of the creators' motivation in choosing a particular brand name. As a matter of fact, some of the most creative minds are engaged in brand name creation. Brand names have to fulfill a number of legal and commercial requirements. The legal constraints are the various copyright and trademark laws, which tend to exclude offensive words as well as words of public interest (e.g. *federal*, *national*) and disallow words that are similar to brand names already in use. From a commercial point of view, brand names have to be distinctive and memorable: they have to aid the advertising of a product and encapsulate the brand's image. Creating a name that fulfills all these functions may prove to be very difficult, particularly in today's global marketplace, where a name should ideally accomplish that goal in a number of languages. Despite the extensive market research that goes into brand name creation (in the late 1990s, the creation of an international brand name cost around $60,000), grotesque multilingual failures are not unheard of. The Mitsubishi *Pajero,* for instance, failed in Spanish-speaking markets because *pajero* means 'masturbator' there. Similarly, the Chevrolet *Nova* did not sell well in Spanish-speaking markets because *no va* translates as 'does not move'—obviously a poor omen for a car.

There can be no doubt that the relationship between a name and its referent has more psychological reality for language users than that pertaining between any other linguistic signs and their referents. This observation has excited philosophers for centuries, as is evidenced in Plato's *Cratylus*. Modern linguistics, philosophy, and onomastics, however, have tended to dismiss this widely held assumption as 'primitive' or 'superstitious'. This stance has led to an ever-widening chasm between limited academic interest in names and naming (as evidenced e.g. in the small number of academic programs in onomastics) and widespread popular interest (as evidenced e.g. in the innumerable internet sites devoted to names and naming). It seems that Juliet's question remains as enigmatic as ever: 'What's in a name?'

References

Abate, Frank R. (ed.) 1994. Proper names master index: a comprehensive index of more than 200,000 proper names that appear as entries in standard reference works. Detroit: Omnigraphics.

Algeo, John. 1973. On defining the proper name. Gainesville: University of Florida Press.

Arney, Roberta. 2001. Onomastics bibliography. El Paso: University of Texas at El Paso Library. <http://libraryweb.utep.edu/ref/onomast.html>

Baker, Ronald L. 1991. The study of place names. Terre Haute, IN: Indiana Council of Teachers of English, Hoosier Folklore Society.

Dickson, Paul. 1996. What's in a name? Reflections of an irrepressible name collector. Springfield, MA: Merriam Webster.

Eichler, Ernst, Gerold Hilty, Heinrich Löffler, Hugo Steger, and Ladislav Zgusta (eds.) 1995/1996. Namenforschung/Name Studies/Les Noms Propres: Ein Internationales Handbuch zur Onomastik/An international handbook of onomastics/Manuel International d'Onomastique, 2 vols. Berlin, New York: Mouton de Gruyter.

Frazer, James. 1993. The golden bough: a study in magic and religion. Ware, UK: Wordsworth Reference; 1st edition 1922.

Koss, Gerhard. 1990. Namenforschung: eine Einführung in die Onomastik. Tübingen: Niemeyer.

Matthews, C.M. 1972. Place names of the English-speaking World. New York: Scribners.

Morgan, Jane, Christopher O'Neill, and Rom Harré. 1979. Nicknames. London: Routledge and Kegan Paul.

Piller, Ingrid. 1996. American automobile names. Essen: Blaue Eule.

Smith, Elsdon C. 1973. The new dictionary of American family names. New York: Harper & Row.

A useful gateway to naming on the internet is provided at http://jerryhill.tripod.com/

INGRID PILLER

See also **Etymology; Proper Nouns; Reference**

Natural Classes

The notion of natural class is widely used in phonological theory for the categorization of segments or phonemes, our focus here, although the notion can also be applied to other domains—for example, a group of morphemes may constitute a natural class. As we will see, this notion is useful for descriptive

purposes, but plays no direct role in grammatical theory. Given this derivative status, we begin with an informal example, instead of choosing one of the various formalizations of the notion that are implicitly or explicitly used in the literature.

Informally, the featural representation [voiceless stop] defines a natural class of English phonology, which includes the sounds [p, t, k] and no others. The other sounds of English are either voiced (e.g. [a, m, g]) or not stops (e.g. [f s]). Because natural classes are defined in terms of a conjunction of shared properties, it is generally the case that using *more* features, yielding a more specific (richer) specification, will define a class with *fewer* members: the class defined by [voiceless stop] has fewer members than the class defined by [stop], since the latter includes voiced and voiceless stops. Several issues arise when discussing natural classes: (i) language specificity, (ii) organization of the feature inventory, (iii) underspecification, (iv) relation to substantive (e.g. phonetic) categories, (v) level of analysis (UR, SR), and (vi) status in grammatical theory. We briefly consider each of these.

(i) *Language specificity*: It is crucial to note that natural classes can only be defined with respect to a given language. For example, in English, [t, n] do not form a natural class of sounds: using binary valued features, their representations both contain the set of features [−continuant, + coronal]. However, this set of features is also contained in the representation of [d], which also occurs in English. Any description of [t,n] that merely conjoins features cannot include [t, n] but excludes [d]. However, in a language where there is no [d], where say [t] and [n] are the only noncontinuant coronal consonants, [t, n] would constitute a natural class of sounds.

(ii) *Organization of feature inventory*: If phonological representations are assumed to consist not merely of unordered sets of features, then natural classes cannot be defined in simple set theoretical terms—members of a set are not organized in any way. For example, in theories of *Feature geometry,* which impose hierarchical groupings of features, the more complex notion of subsumption must be used.

(iii) *Underspecification*: Another formal problem arises with the use of certain kinds of underspecified representations. If, at a given grammatical level, two representations *A* and *B* differ in that *A* contains a subset of the information contained in *B*, then it is not possible to distinguish the representation of the natural class containing both *A* and *B* from the representation of just *A*. In other words, *A* and *B* are not distinct in the sense of Chomsky and Halle (1968:328). This issue arises in models that treat certain segments as defaults. If [t] is treated as a default stop, with no features for place of articulation, then the representation

of [t] ([− voiced, −continuant]) will be a subset of the representation of [p] ([− voiced, −continuant, + labial]) or [k] ([− voiced, −continuant, + velar]). In such a system, the representation of [t] is identical to the representation of the class [p,t,k].

(iv) *Relation to substantive categories*: Does every formally definable natural class constitute a natural phonetic (acoustic or articulatory) grouping of sounds? Clearly, there is a fairly straightforward match between a natural class such as the voiceless fricatives [−voiced, + continuant] of English and a cluster of phonetic properties. However, it is less clear that a definable natural class such as [−lateral], which would potentially include all the sounds of English except for the lateral [l], forms a phonetically identifiable grouping. There are several approaches to take. First, it may not be the case that [−lateral] describes all the sounds of English besides [l]. Perhaps the vowels are completely unspecified for the feature [lateral]; perhaps the feature [lateral] is only present on the non-nasal sonorants [l] and [r], the first being [+ lateral] and the second [−lateral]. Second, just because a natural class is definable formally, it does not follow that it must play a role in the phonology of every language, or even of any language.

The natural classes that can be defined depend upon the set of features available to the language faculty; thus, the set of natural classes is theory dependent. For example, given a set of features that distinguishes the vowels [i, I, e, E] on the basis of height alone, there is no way to group [i, e] or [I, E] into natural classes by themselves. However, if the feature system represents the former pair as [+ tense] and the latter as [−tense], we can represent the relevant classes. Phonetic factors ([+ tense] vowels share articulatory properties) and phonological factors such as the existence of harmony systems that treat [i, e] uniformly as opposed to [I, E] support the existence of the relevant feature.

This brings us to an apparent circularity of reasoning: rules or processes are *defined* as applying to natural classes of sounds, and natural classes are defined on the basis of features that are posited when a set of sounds is treated uniformly by rules. For example, in Lamba, there is a process that palatalizes [s] to [S] before [i]. There is also a palatalization of [k] to [C] before [i]. It is tempting to collapse these processes into a single rule; however, there is no feature system that will treat [s, k] as a natural class to the exclusion of [t], which is also found in the language and which does not undergo palatalization. Should we take the Lamba data as evidence of a new feature [F], such that [s, k] are [+ F] and [t] is [−F]? There are several reasons not to do so. First, our identification of the two processes may be somewhat superficial—palatalization' is a descriptive term that does not obviously rep-

resent a single process. It covers a wide range of synchronic and diachronic phenomena. Second, we can restrict the distinctive feature theory by requiring that every featural distinction be transducible from the acoustic signal. There is no possible acoustic parameter that is shared by [s, k] to the exclusion of [t]. In other words, we can take a phonetic parameter to be a necessary (but not sufficient) correlate of a distinctive feature.

(v) *Level of analysis*: It will have been noted that allophonic variation was not taken into account in the discussion above. For example, it was stated that [p, t, k] constitute the natural class of voiceless stops in English, without acknowledging the existence of aspirated stops [pʰ, tʰ, kʰ]. Should natural classes be defined at the level of Underlying Representation, the level of Surface Representation, or at both these levels? Since, as the following paragraph will argue, the notion of natural class is a mere descriptive convenience, natural classes can be referred to wherever it is convenient to do so, as their existence is epiphenomenal.

(vi) *Status in linguistic theory*: Given the dubious status of the phoneme in generative phonology, and the equally difficult notion of segment, the question arises of the status of the notion natural class. Is the notion useless, a useful descriptive device without real theoretical content, or an actual element of linguistic theory? The best view is probably the second—a natural class is just a convenient way to refer to the set of representations a rule applies to, but speakers' knowledge of phonology need not contain the notion. Natural classes of segments are affected by rules since the Structural Description of the target of a rule may contain a subset of the information found in any particular token of a representation that the rule affects. For example, if a grammar contains a rule

$$[-\text{sonorant}] > [-\text{voiced}] / \_\_ [-\text{voiced}]$$

then any representation that is [−sonorant] will undergo the rule in the right context, but this set of representations need not be identified within the grammar as constituting a natural class, independently of this rule.

References

Chomsky, Noam, and Morris Halle. 1968. The sound pattern of English. New York: Harper & Row.

Clements, G.N., and Elizabeth Hume. 1995. The internal organization of speech sounds. The handbook of phonological theory, ed. by John Goldsmith, 245–306. Oxford: Blackwell.

CHARLES REISS

See also **Feature Theory**

Navajo and Athabaskan-Eyak Language

The Navajo language is one of several Southern Athabaskan languages, more commonly known as Apachean languages, spoken in the Southwest. The Athabaskan languages, along with Tlingit and Eyak, make up the larger Na Dene language family. There are three major geographical groupings of Athabaskan languages:

(1) Northern Athabaskan—including about 25 languages spoken in Alaska and Canada, including Koyukon, Holikachuk, Tanacross, Ahtna, Dena'ina, Deg Hit'an, Tanana, Upper Kuskokwim in Alaska, Han and Gwitch'in in Alaska and Canada, Tagish, Tutchone, and Tahltan at the Yukon headwater, Tahltahn, Sekani, Tsetsaut, Babine, Carrier, and Chilcotin in the far west of Canada, and Hare, Bearlake, Mountain, Kaska, Dogrib, Slave, Beaver, Chipewyan, and Sarcee across the rest of western Canada. It is important to note that, rather than discrete groups, the Alaskan and Canadian Athabaskan languages constitute a language and dialect continuum, i.e. neighboring languages/dialects may be mutually intelligible.

(2) Pacific Coast Athabaskan (and extinct Kwalhioqua-Tlatskanai)—including Hupa and Tolowa in California, and Tututni in Oregon.

(3) Navajo and Apache—including seven Apachean languages divided into two main groups: (1) the Western group, including Navajo in one branch and a second branch divided into the San Carlos group (San Carlos, White Mountain, Cibecue, etc.) and the Mescalero-Chiricahua group, and (2) the Eastern group, divided into a branch containing Jicarilla and Lipan and a second branch containing Plains Apache.

Of the Athabaskan languages, all but Navajo are moribund, with little if any intergenerational transmission. Even Navajo, with approximately 125,000 speakers, is threatened: while more than 95% of children entering bilingual education programs were fluent in Navajo in the mid-1970s, today fewer than 50% of

kindergartners are fluent. Of the Apachean languages, Navajo is by far the best documented, and indeed may be the best documented of any Native American language. This is due in large measure to the pioneering work of Robert Young and William Morgan. Their grammars and dictionaries of Navajo over the past half century (e.g. 1987, 1992) have set the standard for work both in Apachean and in Athabaskan more generally.

Phonology

The distinctive sounds (phonemes) of Athabaskan include obstruents, sonorants, and vowels. The ancestor language—Proto Athabaskan—likely contained the following stop obstruents:

T   TL   TS   CH   K^y CH^w Q

The modern languages show a range of shifts among the Proto-Athabaskan stops. The following list shows the modern reflexes of the Proto-Athabaskan word *-tsi'* 'head':

Tanaina	*-tsi'*
Tanana, Han	*-tthì'*
Koyukon	*-tli'*
Gwitch'in (AK)	*-kì'*
Gwitch'in (Can.)	*-chì'*
Hare	*-pfí', fí'*
Bearlake	*-kwí', kfí'*
Navajo	*-tsi:'*

The sonorants of Proto Athabaskan were *w* and *y*, *m, n*, and a palato-velar nasal. Proto Athabaskan vowels are reconstructed as *i, e, a, u* plus three reduced vowels (mid, low, and back). All of the modern Athabaskan languages have both oral and nasal vowels. Some of the languages have short and long vowels; others have full and reduced vowels.

The phonemes of Navajo include plain, aspirated, and glottalized stops and affricates (*d, dl, dz, j,* and *g* are the symbols used for the plain stops), fricatives (*s, z, sh, zh, h, gh*), and liquids and sonorants (*l, ł, y, w, n, m*). Navajo has plain and nasalized short and long vowels, and high vs. low tone (distinctive pitch).

Morphology and Verb Structure

The Verb Complex

The Athabaskan verb is a complex polysynthetic structure made up of a stem plus prefixes. The stem is composed of a root plus suffixes (or other modification) indicating mode and aspect. Prefixes indicate subject and objects, mode and aspect, and adverbials. The prefix complex can be analyzed as a template comprised of basic positions or zones preceding the stem. The prefix chart in Figure 1 gives a general idea of the ordering of prefixes. Note that the stem is in position zero and subject prefixes are in positions 2 and 5, with tense, aspect, and mode in 3, and direct object in 6.

Some of the prefixes noted above are obligatory in the sense that the particular position they occupy must be filled by one of the possible variants of that prefix class in every derivative of the verb. These obligatory prefixes are the person, tense, aspect, mode, and voice/valence (also called the 'classifier') prefixes. Other, derivational, prefixes, e.g. adverbial, iterative, are optional—he positions these prefixes occupy may or may not be filled depending on the meaning of the derivative. Sometimes, prefixes are thematic—they encode a situational participant—and they are lexicalized, which means that they are present in every derivation of a particular verb. These thematic prefixes in combination with the voice/valence prefix and the root of the verb are referred to as the VERB THEME.

The Athabaskan theme is the underlying skeleton or verb construction to which prefixes and suffixal elements are added in producing an utterance. The theme itself has a meaning and is the basic unit of the Athabaskan verbal lexicon. Verb forms derived from themes have, in addition to the stem (i.e. the aspectually suffixed root) and theme prefixes, inflectional prefixes (person, number, etc.) and derivational prefixes (aspect, adverbials, etc.). These prefixes are arranged in relatively fixed positions preceding the stem as shown in Figure 1.

In Navajo, the THEME consists of the valence prefix (position 1) plus the stem (position 0). The VERB BASE consists of the mode—conjugation prefixes (position 3) plus the subject pronoun (position 2) plus the verb theme. Examples of verb themes in Navajo, with verb words derived from them, are shown below:

Theme:	*na + θ + né*	'play'
Example verb:	*naashné*	'I'm playing'
Analysis:	*na + θ + sh + né*	
	Thematic + imperf + 1st person subject + stem	
Theme:	**ha # O + ł + géésh**	'cut O (object) out'

10	9	8	7	# #	6	5	4	3	2	1	0
post positions	derivational/ thematic	reflexive reversionary semiliterative iterative	distrib./pl	disjunct boundary	DO	Subj. (3rd)	derivational/ thematic	mode/ conjugation	Subj. (1st & 2nd sg & pl)	voice/valence	stem

Figure 1. Navajo prefix chart.

Example verb: ***hadeiilgéésh*** 'we're cutting it out'
Analysis: ***ha # da # θ + θ + sh + ł + géésh***
Thematic # distributive plural # 3rd person direct obj. + imperf. + 1st person subj. + valence + stem

The Aspectual System

Aspect in Athabaskan is expressed by a complex, multidimensional system. There are two major categories that combine in expressing the temporal contour of the state or activity described by the verb. The convention in Athabaskan linguistics is to call one of these two categories MODE and the other ASPECT (although some researchers refer to these categories as 'aspect' and 'aktionsart', respectively). Both of these categories are obligatory—every verb must be marked for both mode and aspect. This morphological marking is an intersecting one, i.e. mode and aspect are marked by prefixal and suffixal elements, which encode both categories.

There are seven modes in Navajo, describing activity as incomplete, complete, ongoing, future, potential, customary, or recurrent. Examples of verbs in the imperfective and perfective modes are:

Imperfective: ***yáshti'*** 'I'm talking'
Perfective: ***yáłáti'*** 'I talked

The second basic temporal category is aspect, which describes the manner in which an activity or event is carried out over time—whether it happens once, or repeatedly, or at length. There are 12 aspects in Navajo and they allow the expression of such distinctions in meaning as: 'I am red' vs. "I turned red' and 'I am chewing it' vs. 'I bit it'. For example:

Durative: **yáshti'** 'I'm talking'
Momentaneous: **'ayániishtééh** 'I'm starting to talk"
Repetitive: **yádíshtih** 'I'm chattering'

Classificatory Verbs

Athabaskan verbs in general show a wide range of classificatory functions. Some verb stems are specialized according to whether the subject of the action is singular (or dual) or plural. Others are specialized for singular vs. plural objects. Still other actions are referred to by different stems depending on the physical characteristics of the subject or object involved. These stems are the classificatory verb stems. The Athabaskan classificatory verbs categorize the material, shape, consistency, size, animacy, arrangement, quanta, and containment of the subjects of intransitive verbs of position and location, and the objects of transitive verbs of handling.

The 11 Navajo classificatory verbs delineate (1) solid, round, or compact objects, (2) slender, flexible, or paired objects, (3) slender, rigid, or stick-like objects, (4) animate objects, (5) plural separable objects, (6) plural objects in profusion, (7) noncompact matter, (8) objects in open containers, (9) flat, flexible, or cloth-like objects, (10) mushy wet objects, and (11) heavy, large, or bulky objects.

Word Order

The word order of the Navajo sentence is Subject–Object–Verb (SOV).

Łééchaa'í mósí yinoołchééł
Dog   cat   it's chasing it
S   O   V

Most transitive Navajo sentences do not name both the subject and object in noun phrases. When only one is mentioned as a noun phrase, it is usually the direct object, that is, the new information. In this case, the direct object has an agreement prefix on the verb, which is ***yi-***. However, when the topic, or old information, is mentioned in a noun phrase, this unusual case is marked by a different object agreement prefix, ***bi-***. For example:

Mary hayííłtî 'He carried Mary up'
Mary habííltî 'Mary carried him up'

The Future of Athabaskan Languages

Work on language revitalization is under way in many Athabaskan-speaking communities, and efforts in the Navajo community are perhaps the most robust. There are language programs in many schools across the Reservation, and at the two branches of Diné College. The language is also taught at the state universities of New Mexico and Arizona. The Rock Point School on the Navajo reservation is a model bilingual education program. At Rock Point, children come expecting that they will succeed and they do—the Navajo program gives students pride in being Navajo, in their language, and in their culture. This effective program includes lessons in Navajo by community elders on Navajo cultural matters and a strong involvement of parents and so family both in conferences and in activities such as Language Fairs and book-making nights.

References

Axelrod, Melissa. 2003. The semantics of time. Lincoln: University of Nebraska Press.

Cook, E-D., and Keren Rice (eds.) 1989. Athapaskan linguistics: current perspectives on a language family. Berlin: Mouton.

Faltz, Leonard. 1998. The Navajo verb: a grammar for students and scholars. Albuquerque: University of New Mexico Press.

Fernald, Ted, and Paul Platero (eds.) 2000. Athabaskan syntax and semantics. Oxford: Oxford University Press.

Jelinek, Eloise , Sally Midgette, Keren Rice, and Leslie Saxon (eds.) 1996. Athabaskan language studies: essays in honor of Robert W. Young. Albuquerque: University of New Mexico Press.

Kari, James. 1979. Athabaskan verb theme categories. Fairbanks, AK: Alaska Native Language Center Research Papers, #2.

Leer, Jeff. 1979. Proto-Athabaskan verb stem variation, part one: phonology. Fairbanks, AK: Alaska Native Language Center Research Papers, #1.

Rice, Keren. 2000. Morpheme order and semantic scope: word formation in the Athapaskan verb. Cambridge: Cambridge University Press.

Sapir, Edward, and Harry Hoijer. 1967. The phonology and morphology of the Navaho language. Berkeley: University of California Press.

Young, Robert W. 2000. The Navajo verb system: an overview. Albuquerque: University of New Mexico Press.

MELISSA AXELROD

Neurolinguistics

Neurolinguistics is a relatively new discipline that comprises both clinical and basic research into the functional relationship between language and the brain. The research topics of neurolinguistics require a multidisciplinary approach and share parts of traditional disciplines such as linguistics, medicine, psychology, and computer science, and more precisely, parts of the fields of psycholinguistics, aphasiology, neurology, brain imaging, neuropsychology, and neuroinformatics. Due to the combination of theories and methods, all these fields contribute to a new discipline in life sciences, called cognitive neurosciences. As for linguistics, this leads to an innovative research area providing empirical and neurophysiological methods for testing aspects of linguistic theories and models.

Looking back on the first emergence of common research methods in the history of linguistics, five main stages can be observed until now: (1) language philosophy, using introspection with an over 2000-year-old tradition; (2) first empirical studies on genealogy and roots of languages, using comparative approaches in typology for more than 200 years; (3) empirical behavioral studies of psycholinguistics, using e.g. reaction time experiments, with an almost 50-year-old tradition; (4) computer-aided simulation of language processes in computerlinguistics, using e.g. neural networks, for more than 30 years; (5) brain imaging of language processing and higher brain function in cognitive neurosciences during the last 20 years.

Therefore, neurolinguistics is one of the newest, but also one of the most rapidly growing fields in contemporary linguistics. At present, neurolinguistics can be divided into three major branches, depending on the research domain (see Figure 1).

Clinical Neurolinguistics or Clinical Linguistics

This field, formerly also called patholinguistics or aphasiology, mainly deals with clinical aspects of brain-related language disorders (aphasia). Aphasia is an acquired language disorder caused by brain damage, which may induce severe impairment in both language comprehension and production, as well as in reading and writing. Reasons for brain damage are mainly strokes, due to an interruption of the blood flow that causes the death of nerve cells in the respective area. Possible reasons for the interruption may be a blocked vessel (arteriosclerosis), a blood clot (thrombosis), or a burst blood vessel accompanied by bleeding, e.g. caused by a ballooning expansion of a weak vessel wall (aneurysm). Moreover, traumatic head injuries and brain tumors in certain areas may cause such language deficits. Types of aphasic language disorders range from minor difficulties, which concern only naming, to major difficulties such as a complete loss of language. The strength of the impairment depends on the location and extent of the brain damage. Currently, aphasia affects more than 1,000,000 individuals (stroke survivors) in the United States. By combining linguistic theory and neurological knowledge on aphasia, clinical neurolinguists typically work in stroke units and rehabilitation clinics, providing mainly three occupations: (1) Diagnosis: In order to test aphasic patients on the extent and the special type of language deficit in standardized examinations, several neurolinguistic test batteries are used: e.g. the *Boston Diagnostic Aphasia Examination*
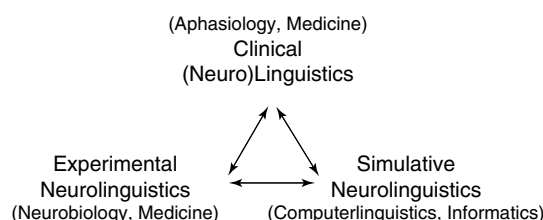
Figure 1

(BDAE), *Porch Index of Communicative Ability* (PICA), or the *Western Aphasia Battery* (WAB). (2) Speech therapy: This represents the main activity of the clinical neurolinguists, who have to perform adapted speech therapy programs with the patients for several weeks or months. (3) Clinical research: This part of their occupation covers the linguistic aspects of aphasia research in hospitals and universities and results in the optimization of treatment efficacy and effectiveness of therapy programs.

Experimental Neurolinguistics

In contrast to clinical neurolinguistics, this field mainly covers basic research on the neurobiology of language of healthy probands. Experimental neurolinguists try to understand the nature of representation and physiological processes that contribute to normal language by investigating underlying neural mechanisms in general. Their main goal is to study the neurophysiological phenomena in the brain during language comprehension and production. Combining state-of-the-art neurophysiological noninvasive techniques with linguistic, neurobiological, and neuropsychological findings, experimental neurolinguists investigate certain aspects of language processing under laboratory conditions. Since this work requires high-tech equipment and a broad expertise in several disciplines, this kind of research involves typical teamwork, often represented by more than one laboratory.

Common noninvasive techniques for the study of language processes in healthy participants are the following techniques, which represent the most frequently used empirical methods in neurolinguistics. Some invasive techniques, only used during preoperative diagnostics of patients, are described later.

(1) Electroencephalography. By using scalp disc-electrodes, which are attached on to the head skin by an electrolyte conductive gel, human electroencephalography (EEG) is a real noninvasive technique, first described in 1929 by Hans Berger. This technique gained its current importance in neurolinguistics after powerful computers became available. Recording the electrical activities of the underlying neural substrate (at least several 10,000 neurons), the EEG signal provides information on the time course of certain cognitive processes. Since the EEG measures the brain's activity directly, it has a very high time resolution below 1 millisecond and this allows assessing 'when' a certain event is processed in the brain. Analyzing the brain waves elicited of a certain event (e.g. onset of a given word), the amplitude and time course of the wave can be interpreted (= event-related potential, ERP). Analyzing the coherence changes within certain frequency bands, the correlated activity of different brain areas during a task can be measured (= coherence analysis). This latter analysis allows to monitor the cooperation of different neuronal networks during cognitive tasks such as language processing.

(2) Magnetoencephalography. Like EEG, magnetoencephalography (MEG) is the second real noninvasive technique, and provides the same advantages of direct measurement of brain processes. In contrast to EEG, MEG measures the magnetic component of electrical activity. By using superconducting quantum interference devices (SQUIDs), which are contactlessly positioned around the head in a so-called Dewar, MEG records the very low magnetic fields accompanying the neural activity. Like the EEG technique, ERP- and coherence analysis of the MEG signal allows investigating language processes in the time and frequency domain.

(3) Positron emission tomography. Even though the positron emission tomography (PET) technique is not based on invasion of the skull, it is a 'small'-invasive technique, since radioisotopes are injected or inhaled. In clinical use since the late 1970s, PET scanners visualize differences in metabolic processes of the brains during a stimulus task (e.g. listening to words) and during a control task (e.g. listening to noise). Short-lived radioactive isotopes are intravenously injected and can be tracked from outside the head with a tomographic scintillation counter after their radioactive decay. While passing the brain tissue, the emitted positron of the former isotope collides after 2 to 8 mm with a body's electron, producing two gamma rays. By detecting and analyzing such coincident gamma rays, the spatial position of the former isotope can be determined. The spatial distribution of the isotopes in the brain allows an insight into the extent of the metabolic processes, and thus the brain's activity during a time window of several seconds. After the computational visualization of the data, false color pictures indicate the brain areas of higher or lower neural activity during a certain language task compared with a control task.

(4) Magnetic resonance tomography (MRT) or *magnetic resonance imaging* (MRI) are synonymously used for a powerful technique, which allows three-dimensional high-resolution brain imaging. After the head is exposed to the strong magnetic field of a cylindrical superconducting magnet, the spin axes of every hydrogen atom of the head brought into alignment. Then, short radio waves of the hydrogen resonance frequency (125 MHz) are sent to the tissue and by absorbing the energy the spin axes of the atoms are disordered for a short moment. After this radio wave stimulation, the hydrogen spin axes return to their

original, randomized state and emit energy in the form of a very weak radio wave. These remitted radio waves deliver the information on the head's structure. After being analyzed and visualized by computer, cross-sectional images of the brain with a resolution of less than 1 millimeter can be made. Outside of clinical use, MRI scanners of up to 11 Tesla are used.

This technique provides high-resolution information on the anatomy of the brain and can be combined with electrophysiological findings of EEG or MEG. In order to monitor cognitive processes in terms of metabolic changes, functional MRI must also be performed. Functional magnetic resonance imaging (fMRI) maps changes in oxygen concentration that correspond to nerve cell activity. By overlaying a map of oxygen concentration changes on the anatomical data, those brain areas that show increasing activity correlated to a given stimulus can be defined. Even though the time resolution is not excellent, the spatial resolution is in the range of several millimeters, which makes fMRI one of the most powerful techniques in neurolinguistic research.

(5) Invasive techniques. During preoperative diagnostics, e.g. prior to epilepsy surgery, electrode arrays are implanted and monitored for several days to identify the focus of abnormal electrical activity. Such intracranial electrodes, placed directly onto the cortex, allow the recording of underlying cortical activity during cognitive tasks, e.g. object naming, speaking, reading, etc. Furthermore, the underlying cortex area can be electrically stimulated by short pulse trains of weak currents. If the stimulated cortex area contributes to language, speech arrest or another aphasiological symptom will occur. In other words, by electrical stimulation, temporary lesions of the brain can be made in order to test the possible role of a given cortex region in the language process. This test is required prior to the performance of a neurosurgical epilepsy operation to assess where certain cerebral functions are centered such as memory and speech.

Another invasive technique applied for the same clinical reasons as mentioned above is the Wada test or *Intracarotid Sodium Amytal Test* (first described by Juhn A. Wada). After the injection of an anesthetic (e.g. sodium amytal) into the right or left internal carotid artery, the respective hemisphere is anesthetized for a few minutes. The type and extent of cerebral dominance of a patient can be assessed during the following behavioral test of language comprehension, production, object naming, and memory abilities. In a healthy, typical right-handed volunteer, suspension of the left (language-dominant) hemisphere leads to a Wernicke aphasia with a neologistic jargon, whereas suspension of the right hemisphere leads to a dysprosody.

Simulative Neurolinguistics

In simulative neurolinguistics, the knowledge of theoretical linguistics, clinical linguistics, neuropsychological case studies, and results of experimental neurolinguistics are used as input for a computer simulation with special software environments. By using computational techniques of computerlinguistics and neuroinformatics (e.g. connectionist neural networks), previously observed real language processes can be used for a computer simulation. During simulation, each condition can be modified step by step, and the results can be obtained immediately. The time required depends only on the computing power of the hardware used. The advantage of this technique is the enormous flexibility of the simulated paradigm. Contrary to real-life processes, where investigators have to wait for patients with certain impairments, virtual impairment can be created and tested within seconds. On the other hand, the reliability of the predictions of such simulations depends on the implemented processes, models, and constraints. However, even though simulative neurolinguistics seem to have an enormous potential for future research, it did not have a remarkable influence on neurolinguistics or cognitive neuroscience until now.

Based on the large progress in brain imaging and computerized analysis of brain data during the last few years, it can be expected that neurolinguistics will strongly contribute to linguistic theory and modeling during the next decades. For a few years now, brain processes during language can be investigated with a resolution in time of less than 1 millisecond and in space of less than 1 millimeter. Thus, a powerful new tool is available for linguists in the at least 2000-year-old attempt to understand human language. All methods developed and used in linguistics (see above) form an emergent instrument for the very different topics, questions, and approaches in studying language within the different fields of linguistics. For most of them, an integration of neurophysiological techniques would be valuable. For example, neurolinguistic studies indicate how and when language is learned as first or second language in a mono- or bilingual child, how different languages are learned by a juvenile or adolescent, and how it is represented in the brain as well as how learning can be facilitated or disturbed. In addition, the contemporary theories and models of language may be improved by neurolinguistic findings. The enormous impact of the behavioral data of psycholinguistics on linguistic modeling since the 1950s did show the necessity and usefulness of empirical evaluation. Neurolinguistic research offers an even stronger and more direct observation of cognitive phenomena and language processes in the brain.

References

Benson, D. Frank, and Alfredo Ardila (eds.) 1996. Aphasia: a clinical perspective. Oxford: Oxford University Press.

Brown, Colin M., and Peter Hagoort (eds.) 1999. The neurocognition of language. Oxford: Oxford University Press.

Caplan, David. 1992. Language: structure, processing, and disorders. Cambridge: MIT Press.

Eling, Paul (ed.) 1994. Reader in the history of aphasia: from Gall to Geschwind. Amsterdam: John Benjamins.

Frackowiak, Richard S.J., Karl J. Friston, Christopher Frith, and Raymond Dolan (eds.) 1997. Human brain function. San Diego: Academic Press.

Friederici, Angela D. (ed.) 1999. Language comprehension: a biological perspective. Berlin: Springer-Verlag.

Gazzaniga, Michael S. (ed.) 2000. The new cognitive neurosciences. Cambridge: MIT Press.

Grodzinsky, Yosef, Lew Shapiro, and David Swinney (eds.) 2000. Language and the brain: representation and processing. San Diego: Academic Press.

Kertesz, Andrew (ed.) 1997. Localization and neuroimaging in neuropsychology. San Diego: Academic Press.

Rugg, Michael D. (ed.) 1997. Cognitive neuroscience. Cambridge: MIT Press.

Stemmer, Brigitte, and Harry A. Whitaker (eds.) 1998. Handbook of neurolinguistics. San Diego: Academic Press.

Zemlin, Willard R. 1998. Speech and hearing science: anatomy and physiology, 4th edition. Boston: Allyn & Bacon.

HORST M. MÜLLER

See also **Clinical Linguistics; Localization of Linguistic Information**

New Guinea

New Guinea, the second largest island in the world, lies roughly 100 miles north of Australia. It contains more than 1,000 languages, about one sixth of the world's total (or one language in every 900 square kilometers), making it linguistically one of the most dense and complex regions of the world. The name was given to the island by the Spanish navigator Ortiz de Retes, possibly because he saw some resemblance between the indigenous population and the inhabitants of the Guinea coast of Africa. Politically, the island is divided between two countries: Papua New Guinea, occupying the eastern half of the island, and some 600 associated islands, the largest of which are New Britain, New Ireland, and Bougainville, and Irian Jaya (West Irian), a former Dutch colony until 1963, now the easternmost province of Indonesia. An independent nation since 1975, Papua New Guinea is historically an amalgamation of what were two separate colonies: Kaiserwilhelmsland, occupying the northeastern half of the mainland and islands under German rule from 1884 until 1914, and subsequently mandated by the United Nations to Australian administration under the name 'New Guinea'; and Papua, the southern half of the mainland, formerly British New Guinea, which was proclaimed the Australian territory of Papua in 1906. Papua New Guinea has the highest population (5.2 million) and land mass of the Pacific island nations, with 64% of its population and 84% of its land mass. Irian Jaya has a population between 1.8 and 4 million and occupies a land mass approximately the size of California.

New Guinea was originally peopled by many different waves of migrants, whose prehistory is largely unknown. Although human habitation of the island extends back some 40,000 years, recorded history is very recent and in some cases goes back only a few decades. It was the last major land area in the world to be colonized by European powers; almost all regions have a history of contact of less than a century. The terrain is extremely rugged, with mountains reaching altitudes of more than 15,000 feet, dense tropical rain forests, and fast-flowing rivers, which have long cut the interior of the country off from outsiders. Many villages have no road or river links with other centers, and some can be reached only by walking for up to two weeks.

The peoples of New Guinea are mainly settled villagers living in a subsistence economy. Their productive activities vary according to the zone they inhabit on the island's extraordinary vertical ecology. The extreme highlands have an alpine climate, with widespread frost. There, an intensive agriculture based on the sweet potato has developed over the last few hundred years. This highly productive system has given rise to a local population boom, and the highlands support large, dense groups with large languages. The ten largest indigenous languages of Papua New Guinea belong to the large groups of the interior Highlands; they have from 30,000 to 100,000 speakers. Between them, they account for nearly one third of the population. Perhaps 80% of the languages have fewer than 5,000 speakers, and as many as one third

have fewer than 500. These languages are found mainly in the coastal lowlands and the intermediate areas known as the highland fringe, consisting of pockets of rain forest, swamps, and grassland, where the population is low and thinly spread.

Linguists generally recognize two major language groupings in New Guinea: Austronesian and non-Austronesian (or Papuan). The Austronesian languages clearly constitute a family with a common ancestor, Proto-Austronesian, and comprise some 450 descendant languages in the Pacific basin and another 600 to 700 outside it, making it the largest language family. About one quarter of New Guinea languages are of Austronesian origin, and most of them (with the exception of the languages of western Irian Jaya) belong to the Oceanic subgroup established by the German linguist Otto Dempwolff. The relationships among the non-Austronesian (Papuan) languages are less clear, and the label is best seen as a cover term for a number, perhaps as many as 60, of distinct families.

Figure 1 shows the distribution of the Austronesian and Papuan languages. Most linguists agree that the coastal distribution of most of the Austronesian languages indicates the later arrival of their speakers. The immediate ancestors of the Proto-Oceanic speakers migrated from eastern Indonesia through eastern Irian Jaya into the Bismarck Archipelago. The interior of New Guinea, where the majority of Papua languages are spoken, experienced no European contact until shortly before (and even in some cases some time after) World War II, and hence most of the languages were unknown to the outside world until quite recently. A number are classified as isolates, i.e. languages that seem to have no known relatives. The Indo-Pacific hypothesis, which attempts to link Papuan languages

with those of Tasmania (but not mainland Australia) and the Andaman Islands in the Indian Ocean, has not been generally accepted, nor has the suggestion of a link between Australian languages and the languages of the Eastern Highlands of Papua New Guinea. It is possible that Australia was settled from New Guinea because they were one continent until around 8,000 years ago. Only after the last Ice Age did sea levels rise to separate them.

New Guinea languages are typologically diverse, displaying many interesting and unusual linguistic features. There are languages with SVO (subject–verb–object), SOV, VSO, VOS, and OSV word orders; OVS is the only unattested word order, and this order is quite rare across the world. Most Papuan languages, however, tend to be verb final. In addition, there are examples of noun classifier systems, such as in the Papuan language Nasioi, with a set of more than 100 suffixes added to nouns, adjectives, numerals, and derived nominals to classify the entity being referred to. Both Papuan and Oceanic languages, however, tend to have relatively simple sound systems, with sound inventories smaller than that of English. Consonant clusters are absent or rare in Austronesian, but they are present in many Papuan languages.

Almost all Oceanic languages distinguish between inclusive (referring to the speaker and addressee or addressees, 'I + you') and exclusive first-person pronouns (referring to the speaker and some other person or persons, 'I + he/she/it/they'), as well as a three-way distinction in number between singular (one), dual (only two), and plural, or paucal (more than two). Some languages have a trial, referring to only three. Very few Oceanic languages mark gender in pronouns, and most have a three-way distinction in their
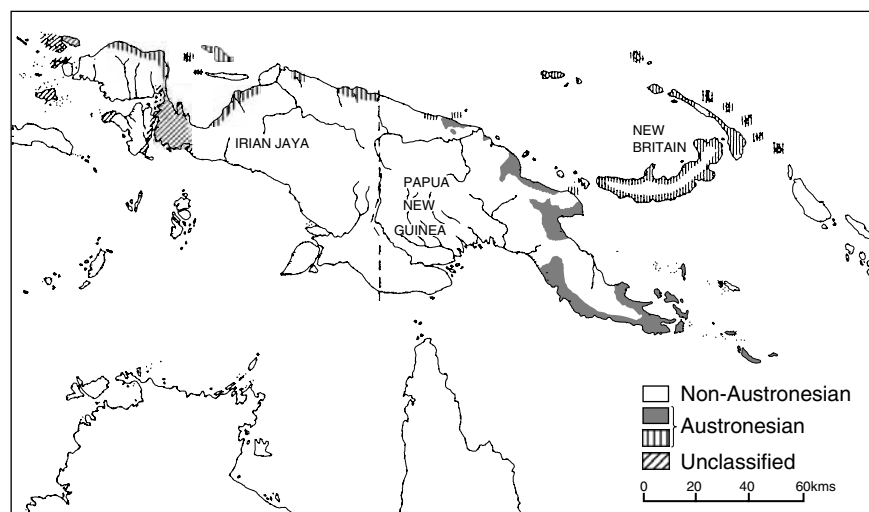
Figure 1. Map of New Guinea showing New Britain and the distribution of Austronesian and non-Austronesian languages.

demonstratives between proximate ('this' near the speaker), intermediate ('that' near the addressee), and distant ('that' away from both speaker and addressee). Most Oceanic languages also distinguish two types of possessive marking, with all nouns being classified into alienable and inalienable. Alienable nouns refer to objects that belong to the speaker, where the speaker typically has control over the possession and may reject it (e.g. food, clothes, etc.). Inalienable nouns refer to items inseparably connected to the speaker, such as relatives or body parts. Inalienable nouns are marked with suffixes that vary in a minor way from language to language.

Only a small number of Oceanic languages have a contrast between active and passive. Grammatical distinctions, such as tense, are mostly expressed by independent particles rather than affixes:

Tuna yau ga   gire
     'I'   PAST see
     'I saw it'.

Papuan languages, however, often have a rich and strongly developed affixal word structure:

Yimas   ama-wa-t
     'I' go PERFECT
     'I went'

Pronoun systems vary widely in Papuan languages but are generally not as complex as in the Oceanic languages. Very few Papuan languages distinguish inclusive and exclusive first-person pronouns, although a number of them distinguish gender in pronouns. Articles are virtually nonexistent, and possessive constructions are less complex than in Oceanic languages. A number of languages have more than one existential ('be') verb, such as Anggor in the Sepik region of Papua New Guinea, which has 18. The different verb forms depend on the shape of the object, its location, and posture. In Enga, 'there are men' is expressed as 'the man stands', whereas 'there are women' is expressed as 'the woman sits'.

Switch reference, although generally rare in the Pacific, is typical of complex sentences in Papuan languages. When two or more clauses join to form a complex sentence, the last verb in the clause retains the subject-tense marking of the first verb, but the other (medial) verbs do not. Instead, they incorporate a suffix indicating whether the subject of the verb is the same as or different from the subject of the following verb.

Nasioi   kad-o-ma   nan-ant-in
     talk-'I'-SAME ACTOR   go-'I'-IMMEDIATE PAST
     'While talking, I went.'

     da? po-ko   nan-amp-e-ain
     you come-DIFFERENT ACTOR   go-'we'-FUTURE
     'When you come, we two will go.'

Much less research has been done on the approximately 250 languages in Irian Jaya, only four of which are spoken by 40,000 people or more, and somewhat more on the roughly 860 languages in Papua New Guinea. In many cases, missionaries undertook the first studies of languages in New Guinea, and missionary linguist work continues today, particularly in Papua New Guinea, through the Summer Institute of Linguistics.

Contacts between speakers of Papuan and Austronesian languages have led to rapid change and diversification to the point where languages such as Magori and Maisin in Papua New Guinea are difficult to classify as Austronesian or Papuan. The Motu language spoken along the southern coast of Papua New Guinea has the grammatical particle system typical of Austronesian languages, but it has the typical SOV word order of Papuan languages.

tau ese au-na   imea bogarai-na-I vada   e   hado
man a tree-the garden middle-its-at PERFECT he/she/it plant
'The man planted a tree in the middle of the garden.'

The small size of many Melanesian societies has also permitted change to spread more rapidly than in larger societies. Significant changes have affected the basic vocabulary and grammar of the Austronesian language Muyuw, spoken on Woodlark Island in Milne Bay Province of Papua New Guinea, in just a 50-year period. Local vernaculars are seen as a unique badge of identity and distinctiveness. Villagers in one community decided at a meeting that they would be different from other Selepet-speaking villages by adopting a new word (*bunge*) for 'no' to replace their usual word (*bia*) shared by all Selepet speakers.

Multilingualism is widespread, and because people marry outside of their community, husbands and wives often speak different languages. Many people, especially men, know the languages of one or two neighboring communities, or perhaps a language with wider currency around their valley or coastline. Where language groups were large, as in the highlands, only those in the border areas would be multilingual. Where groups were small, everyone was effectively in a border area, and knowledge of multiple languages was universal. In the lowland village of Gapun, whose language Taiap is an isolate spoken by about 80 people, the average number of languages understood by men over 40 was five: the vernacular, a lingua franca, and three or so of the other local languages.

In addition to the indigenous languages, there are a number of pidgins and creoles, as well as languages of the metropolitan powers, particularly English (and formerly also German) in Papua New Guinea and Bahasa Indonesia (and formerly also Dutch) in Irian Jaya. Among the pidgins and creoles are those based on

indigenous languages such as Pidgin Yimas, based on the Papuan language Yimas and spoken in Papua New Guinea's Sepik region, and Hiri Motu ('trade Motu'), based on the Austronesian language Motu, spoken by a quarter of a million people and one of Papua New Guinea's three national languages, along with English and Tok Pisin, an English-based pidgin and creole. Papua New Guinea's most widespread language, Tok Pisin, with over 2 million speakers, is the largest pidgin/creole language in the Pacific. In many parts of Papua New Guinea, children grow up speaking Tok Pisin and no longer acquire their local village language. In Gapun, parents began speaking mainly Tok Pisin to their children, and now children over the age of ten no longer use Taiap. The vernacular languages have very little place in the national life of either Papua New Guinea or Irian Jaya. In the former, English is the main language of government and education, and in the latter, Bahasa Indonesia.

References

Capell, Arthur. 1969. A survey of New Guinea languages. Sydney: Sydney University Press.

Foley, William. 1986. The Papuan languages of New Guinea. Cambridge: Cambridge University Press.

Kulick, Don. 1992. Language shift and cultural reproduction. Socialization, self and syncretism in a Papua New Guinean village. Cambridge: Cambridge University Press.

Lynch, John. 1998. Pacific languages. Honolulu: University of Hawaii Press.

Sankoff, Gillian. 1980. Multilingualism in Papua New Guinea. The social life of language. Philadelphia: University of Pennsylvania Press.

Thurston, William. 1987. Processes of change in the languages of North-Western New Britain. Canberra: Pacific Linguistics Series B, No. 99.

Wurm, Stephen (ed.) 1975. New Guinea area languages and language study, Vol I. Papuan languages and the New Guinea linguistic scene. Canberra: Pacific Linguistics Series C, No. 38.

Wurm, Stephen (ed.) 1976. New Guinea area languages and language study, Vol II. Austronesian languages. Canberra: Pacific Linguistics Series C, No. 39.

Wurm, Stephen (ed.) 1977. New Guinea area languages and language study, Vol III. Language, culture, society and the modern world. Canberra: Pacific Linguistics Series C, No. 40.

Wurm, Stephen (ed.) 1980. New Guinea and neighbouring areas: a sociolinguistic laboratory. The Hague: Mouton.

Wurm, Stephen (ed.) 1982. The Papuan languages of Oceania. Tübingen: Gunther Narr Verlag.

SUZANNE ROMAINE

Nida, Eugene Albert

Eugene Nida's research in linguistics and cultural anthropology spans a period of 60 years, during which he has continuously developed and refined his perception of languages and cultures. His initial studies of classical Greek and the New Testament as well as his early books on morphology and the structure of English were ideal qualifications for dealing with Bible translations, a task that he was asked to fulfill by the American Bible Society. A large part of Nida's works and efforts were indeed devoted to showing translators how to better understand the Bible in order to make it understandable, in their turn, to receptors speaking highly different languages and belonging to most diverse cultures. This activity brought him into contact with over 200 languages and cultures in Asia, Australia, Africa, and the Americas. His views on the way in which languages function are thus based on extensive field surveys.

In his description of the word and sentence structure of languages, Nida has resorted to the concepts of American structural linguistics, placing, however, the meaning at the center of his investigations and replacing formal with referential classes. In his works, he has amply demonstrated that, despite formal similarities, words may develop different semantic relations between themselves and the use of strictly formal criteria of analysis can be misleading. Moreover, meaning cannot be located in words, which have 'fuzzy boundaries', but rather on a 'molecular' level, in their combinations with other words. Context thus becomes a fundamental factor in Nida's analysis of meaning, and he follows Martin Joos (1972) in maximizing its role in lexical combinations.

From word and syntax levels, Nida's research has encompassed, in an increasingly detailed manner, the structures of discourse as well as textual organization, texts being regarded as the basic and ultimate carriers of meaning. Still, his investigation of contexts goes beyond strictly linguistic scrutiny. Like Edward Sapir and Bronislav Malinovski, Nida has constantly emphasized the crucial role of cultures in describing and explaining languages. Moreover, his adoption of a sociosemiotic stance has allowed him to explore the social aspects of language as well as the ways in which

language signs relate to other sign systems and to the real world of referents.

Eugene Nida has made a fundamental contribution to the development of translation theory and practice, and his impact on Bible translators, translation scholars, and translation schools has been outstanding. He has brought rigor and scientific objectivity to the subject by incorporating concepts, methods, and classifications from linguistics, pragmatics, semantics, and discourse analysis as well as necessary data from cultural anthropology, linking languages to cultures and broader contexts of communication. The concept of dynamic/functional equivalence that he has introduced in the translational discourse has shifted the emphasis in translation theory from the faithful reproduction of source messages to the creation of translated texts with a strong communicative impact, focusing on the receivers' needs for clarity as well as on their linguistic and cultural expectations. This receptor orientation gives priority to the content of texts and may entail, in some cases, a more radical formal restructuring of the source text without, however, altering its meaning.

In *The theory and practice of translation* (co-authored with Charles Taber, 1969) and in subsequent books, Nida provides a model for describing the translating process and analyzing meaning in more detail. In his view, the translating process involves analysis, transfer, and restructuring. Analysis consists in a reduction of surface structures to kernels (i.e. substructures), making use of the functional classes of objects, events, abstracts, and relations. At the kernel level, languages are found to 'agree far more than on the level of more elaborate structures' (1969:39), and it is at this level that the transfer into the receptor language occurs. The transformation into a new surface structure through restructuring takes place according to the deemed expectations of the receivers. The model draws on principles used in transformational-generative grammar, although the influence of this linguistic direction on Nida's work has often been overemphasized. For one thing, Nida reverses Noam Chomsky's model by starting from surface structures and moving to kernels and not the other way round. Secondly, Nida is interested in language-in-use (Ferdinand de Saussure's *parole* or Chomsky's performance), and not in the abstract level of Chomsky's deep structures. Thirdly, Nida rejects the idea that languages are strictly rule-governed, as this would leave no space for their creative use. It is nevertheless true that transformational-generative grammar offered him a theoretical perspective for explaining processes of decoding and encoding texts. From a similar stance, he could also account for the similarities between languages on a conceptual and even on a formal level, in keeping with his conviction that effective interlingual communication is always possible, although there is no such thing as absolute communication.

In more recent publications in linguistics and translation theory, Nida undertakes detailed analyses at all language levels, broadening the area of interlingual investigations. The author includes a high number of case studies of Bible translations, of scientific and technical texts, of various literary genres, and of European Union texts, and a multitude of experience-based examples from different languages and cultures illustrating his translation principles.

Debates around Nida's works were mainly generated by ideological and religious speculations on his receptor-oriented position in translation as well as by the principle of 'equivalence of effect', on which dynamic/functional equivalence is based. Although the American scholar has suggested a series of tests for checking and comparing the source to target text receivers' understanding and response, "effect" has been regarded as too vague a notion to serve as a criterion for translation evaluation.

Eugene Nida's approach to linguistics and translation is avowedly eclectic. This encyclopedic perspective enables him to cover all major aspects of languages and their translation. He combines insights from structural linguistics, pragmatics, sociolinguistics, discourse analysis, functional grammar, transformational-generative grammar, semiotics, psycholinguistics, rhetoric, stylistics, information theory, etc. Nida openly declares his mistrust of the holistic systems that 'can stifle creativity and lock minds shut to new evidence' (2003:140). His books have aroused considerable interest among linguists, theologians, and translation scholars, and they have certainly been of great help to all those dealing with intralingual, interlingual, and intercultural communication, be it religious or secular.

Biography

Eugene Albert Nida was born in Oklahoma City, Oklahoma on November 11, 1914. He received his B.A. in Greek and Linguistics from the University of California at Los Angeles an 1936, and an M.A. in Patristics from the University of Southern California in 1939. He taught at the Summer Institute of Linguistics in 1937–1952. He did his Ph.D. on English Syntactic Structures under the supervision of Charles C. Fries, Professor of English and Linguistics at the University of Michigan, in 1943. He was consultant for the American Bible Society and the United Bible Societies, 1943–1981; President of the Linguistic Society of America, 1968: and Translations Research Coordinator for the United Bible Societies, 1970–1980. He has studied languages and cultures, counseling on Bible

translating in more than 90 countries, and has lectured in more than 100 universities, also actively participating in other scientific and academic reunions worldwide. He also received five honorary Ph.D.s. He is author and co-author of more than 40 books and 250 articles. Nida has lived in Brussels since 1995, consulting with Bible societies and translators from the European Union.

References

Black, Matthew, and William Smalley (eds.) 1974. On language, culture and religion: in honor of Eugene A. Nida. The Hague: Mouton.

Fawcett, Peter. 1997. Translation and language. Manchester: St. Jerome.

Joos, Martin. 1972. Semantic axiom number one. Language 48. 257–65.

Larose, Robert.1989. Théories contemporaines de la traduction. Québec: Presses universitaires de Québec.

Longacre, R.E. 1994. Eugene Nida. The encyclopedia of language and linguistics, Vol. 5, ed. by R.E. Ascher and J.M.Y. Simpson. Oxford: Pergamon Press.

Munday, Jeremy. 2001. Introducing translation studies. London, New York: Routledge.

Nida, Eugene Albert. 1946. Morphology. The descriptive analysis of words. Ann Arbour, MI: University of Michigan Press.

———. 1964. Toward a science of translating. Leiden: E. J. Brill.

———. (co-authored with Charles Taber). 1969. The theory and practice of translation. Leiden: E. J. Brill.

———. 1975. Exploring semantic structures. Munich, Germany: Wilhelm Fink.

———. (co-authored with Johannes P. Louw). 1992. Lexical semantics of the Greek New Testament. Atlanta: Scholars Press.

———. 1996. The sociolinguistics of interlingual communication. Bruxelles, Belgium: Les Editions du Hazard.

———. 2001. Contexts in translating. Amsterdam and Philadelphia, PA: Benjamins.

———. 2003. Fascinated by languages. Amsterdam and Philadelphia, PA: Benjamins.

Schmitz, John Robert. 1999. The contribution of Eugene Nida to the theory and practice of translation. Translation-transition. 15th World Congress of FIT, Mons, Vol. 1, ed. by Jean-Marie Vande Walle. Mons: Elma Edities.

RODICA DIMITRIU

Niger-Congo

The Niger-Congo language family, previously labeled *Nigritic*, or *western Nigritic*, is one of the four main language families in Africa. The other language families are Khoisan, Nilo-Saharan, and Afro-Asiatic. Niger-Congo languages are spoken in the basins of the Niger and Congo rivers and in the Nuba Mountains in the Sudan. Joseph Greenberg's (1955) original classification viewed the Niger-Congo family as a sister language of the Kordofanian family. He named the ancestor language from which the two emerged as Niger-Kordofanian. However, Kay Williamson (1989a,b) notes that the Kordofanian family falls within the Niger-Congo family; hence, she uses the label Niger-Congo in place of Niger-Kordofanian for the whole family. The Niger-Congo group of languages is spoken by well over 80% of the population of Africa. Geographically, they are spoken in areas stretching from Senegal (West Africa) to Kenya (East Africa). They also stretch from Sudan (North Africa) to the south in the Republic of South Africa. Niger-Congo has ten daughter languages, namely Adamawa-Ubangi (formerly Adamawa Eastern), Atlantic (formerly West Atlantic), Benue-Congo, Dogon, Gur, Ijoid, Kordofanian, Kru, Kwa, and Mande.

Adamawa-Ubangi

Adamawa-Ubanji, previously called Adamawa-Eastern, is spoken in parts of Central African Republic, eastern Nigeria, northern Cameroon, and southwestern Chad. It has languages like Gbaya, Banda, and Zande. The Adamawa-Ubangi family has two daughter languages: Adamawa and Ubangi. Adamawa has languages like Leko, Duru, Jen, Nimbari, Mbum, Bua, Kim, Day, Waja, Daka, and Fali. Ubangi languages are spoken in the area stretching from northern Cameroon across the Central Africa Republic to parts of southern Sudan and northern Democratic Republic of Congo (formerly Zaire). Ubangi languages include Banda, Gbaya, Ngbaka (Sere, Mba), Sango, and Zande. Sango is a lingua franca of the Central African Republic.

While Adamawa allows words to end in consonants, none of the Ubanji languages, with the exception of Gbaya, allows final consonants. Whereas other language families may have nasal sounds, Adamawa-Ubanji has nasal morphemes (root words). Word order is Subject–Verb–Object, and verb reduplication (doubling of verbs in certain grammatical constructions) is common to all the languages in the subfamily. The

languages have between two and four tonal phonemes, i.e. pitch patterns that may alone distinguish meaning.

Atlantic

Atlantic languages, also called West Atlantic languages, are spoken in West Africa in the area stretching from the Senegal river down into Liberia. Most languages in this family are spoken in Senegal, The Gambia, Guinea, Guinea-Bissau, Liberia, and Sierra Leone. Atlantic has about 45 languages and over 30 million speakers. Languages in this family with large numbers of speakers include Fula, Wolof, Diola, Serer, Manjaku, Balanta, Basari, Limba, Kisi, Sua, and Temne. Important linguistic features include consonant mutation/alternation (specific consonants may change into another in particular sound patterns), noun class, and concord systems in which the choice of specific parts of speech (especially nouns) requires other elements (verbs, adjectives, adverbs, etc.) of the sentence to occur with particular affixes, depending on the grammatical construction.

(New) Benue-Congo

The Benue-Congo language family is the largest sub-family of the Niger-Congo group. It is spoken in the areas in and around the Benue and Congo river basins. It is made up of the former Eastern Kwa—Yoruba, Edo, Nupe, Idoma, Igbo—and Bantoid (subclassified into non-Bantu and Wide Bantu). The New Benue-Congo language family is subclassified into eight sub-families: Defoid (Yoruba, Akoko), Edoid (Edo), Nupoid (Nupe, Ebira, Asu, and Gbagyi), Idomoid (Idoma), Igboid (Igbo), Kainji and Platoid (Kainji, Eloyi, Kagoma, Jukun), Cross River (Obolo, Ogoni, Legbo, Ogbia), and Bantoid (Tiv, Swahili, Kikuyu, Kirundi, Kinyarwanda, Shona, Zulu, and Xhosa). The Bantoid group is by far the largest group in the Benue Congo subfamily. Important linguistic characteristics of this large subfamily include cross-height harmony (where either only tense or only lax vowels occur in words) and vowel coalescence (where two or more vowels merge to become one). Syllable structure ranges from CV (consonant–vowel) to CCV and CVC. Consonant types include such uncommon sounds as breathy voiced plosives /bh dh/ and labio velar sounds /kp gb/. Other important features include noun class and concord systems and an SVO word order. Some of the Bantu languages have a fixed stress pattern—words in Swahili, for example, are always stressed on the penultimate (last but one) syllable.

Dogon

Dogon, which is spoken on the mountains of Mali, was classified as a Gur language until 1989. However, it may also be classified as an independent member of the Volta Congo group. Dogon has six major dialects: Donnɔ Sɔɔ, Tombɔ Sɔɔ, Torɔ Sɔɔ, Jamsay, Togo Kan, and TomO Kan. Important linguistic features include independent nasalized vowels (phonemic nasalization of vowels), vowel harmony (co-occurrence restrictions in the distribution of vowels), a two-tone system, and Subject–Object–Verb word order. In Dogon noun phrases, the noun always comes first.

Gur

Geographically, the Gur group of languages is distributed over a wide area stretching from the Ivory Coast, through Ghana, Togo, Benin to Burkina Faso. Gur languages are also spoken on the fringes of Niger, Benin, and Mali. The Gur phylum subdivides into Oti-Volta, Bwamu, Kurumfe, Grusi, Kirma, Dyan, Gan, Dogoso, among others. There are nearly a hundred languages in the Gur family. Some of the languages in this group include Moore (spoken in Burkina Faso by nearly seven million people), Grusi, Gurenne, Wali (Dagaari), Dagbani, Buli, Kabre, Kotokoli, Kasem, Konni, Tayiri, Kusaal, Bassari, Ntrobo, Sisaala, Waali, Mampruli, and Nafaanra. Over twenty million people speak the Gur languages. Some of the phonological features found in Gur languages are tone, vowel harmony, uncommon consonants such as implosives, coarticulated sounds (e.g. /kp/ /gb/), and a syllable-timed rhythm.

Ijoid

Ijo, the smallest branch of the Niger-Congo language family, is spoken by the Ijo of Nigeria and covers both Ijo and Defaka (Afakani). It is spoken in the Niger River delta region and adjacent riverine areas within the Rivers, Bendel, and Ondo states of Nigeria (Jenewari 1989). The Ijo language family is made up of seven languages: Biseni (Amegi), Okodia, Oruma, Nkoroo, Eastern Ijo (which comprises Kalabari, Okrika, and Ibani), Brass Ijo (Nember-Akassa), and Izon (Bumo, Kolokuma, Mein, and Arogbo). Speakers of Ijo are a little over one million. The Ijo were among the first West Africans to have contact with Europeans, and the Kalabari, an Ijo language, is believed to be one of the first Nigerian languages to be written. An important unique structural feature is a consonant harmony also called 'implosive harmony' (which requires that any given word either contains implosives or plosives, but never both), and a noun class system drawing grammatical distinctions based on animateness and biological gender. The basic word order is Subject–Object–Verb, although different word orders are possible (Object–Subject–Verb and Subject– Verb–Object) if object or subject need to be marked as the topic of conversation.

Kordofanian

The Kordofanian language family is located in the Nuba Mountains of Sudan. Its status in the Niger-Congo family has been widely debated. As mentioned earlier, it was first viewed as a sister family of Niger-Congo in a larger grouping called Niger-Kordofanian, but later it came to be seen as a sub-branch of Niger-Congo. Kordofanian has over half a million speakers and consists of about 20 languages. Moro, Mudo, Talla, Miri, Krongo, Talasa (Tumtum), Tiro, Utoro, Rere, Ngile, Tocho, Goy, Gom (Tegali), and Kalak are some of the languages of this subfamily. Kordofanian languages have dental /d, ð, t/ (tongue touches teeth) and retroflex consonants /ṭ, ḍ/ (tongue curls back). Plosives like /t/ occurring between vowels change to fricatives such as /s/, and the noun class system is reminiscent of those found in other Niger-Congo.

Kru

The Kru language family is spoken in southwestern Ivory Coast and southern Liberia. There are 24 languages in this subfamily and the total number of speakers ranges between three and four million. Some of the Kru languages include Grebo, Klao, Dida, Godie, Bete, Nyabwa, Konobo, Bassa, Gbii, Bakwe, Kuwaa, Aizi, Wobe, Dewoin, SEmE, Guere (Krahn, Wɛɛ), Tepo, Chedepo, and Neyo. Most Kru languages have eight vowel phonemes and a vowel harmony where only vowels from a particular set may occur in any given morpheme. Kru syllables tend to end in a vowel. Central vowels such as /i, u, a,/ are found in some of the Kru languages. Uncommon consonant sounds such as implosives, double articulated sounds, and velar fricatives are found in the Kru languages. Kru languages have subject–verb–object word order with indirect objects preceding direct objects. When there is an auxiliary, then both the direct and indirect objects precede the verb. Sentences may be negated with the help of an auxiliary or a particle as well as changes in tone pattern.

Kwa

Kwa has undergone tremendous reclassification with several languages and language groups such as Ijo, Kru, Yoruba, and others moved to other major language families. The New Kwa, as it is now called, is spoken in Côte d'Ivoire, Ghana, Togo, Benin, and Nigeria by about 30 million people. It subdivides into two main subfamilies, namely Nyo and Left bank. Important subfamilies under Nyo include Avikam-Alladian, Agnéby, Potou-Tano, and Ga-Dangme. Gbe, Avatime-Nyangbo, Kposo, and Keby-Animere are identified as members of the Left Bank subfamily. Important New Kwa languages include Akan, Ewe (and other Gbe languages), Ga, Dangme, Gwa, Avikam, Anyi, Baule, Chakosi, Nzema, Santrokofi, Likpe, Adele, Logba, and Kposo. Important linguistic features associated with some of the New Kwa languages are the occurrence of voiceless and voiced labial and velar fricatives, double articulated sounds (single sounds with two different places of articulations e.g. /kp, tp, gb/), vowel harmony, and tone terracing, where pitch is lowered toward the word end.

Mande

The Mande group of languages is spoken in West African countries: Senegal, The Gambia, Mali, the Ivory Coast, Guinea, Sierra Leone, Liberia, Guinea Bissau, Burkina Faso, Mauritania, Benin, Niger, Nigeria, and Ghana. Bambara, Dyula, Susu, Mende, Kpelle, Vai, Lorma, Loko, Soninke, Kweni, Dan, Maninka, Kpelle, Busa, Bisa, Ligbi, Togo and Bobo are Mande languages. Over 40 languages are found in this family. The Mande languages are spoken by a little over 20 million people. Phonologically, many Mande languages have between seven and nine vowel phonemes, labiovelar stops, consonant mutation, two tones, and tone sandhi (adjacent tones influence one another). Unlike most African languages, Mande languages do not have serial verb constructions. Word order is Subject–Object–Verb. In noun phrase constructions, definite determiners, articles, and plurals tend to follow the noun. However, possessive pronouns precede nouns. Some of the languages such as Vai, Mende, Loma, and Kpelle in this group have combinations of Arabic script, Latin writing systems, and unique African writing systems.

References

Bendor-Samuel, John. 1989. Niger-Congo. The Niger-Congo languages. A classification and description of Africa's largest language family (N-CL). New York: University Press of America.

Bendor-Samuel, Olsen, and White. 1989. Dogon. In John Bendor-Samuel, pp. 169–77.

Dwyer, David. 1989. Mande. In John Bendor-Samuel, pp. 47–66.

Greenberg, Joseph. 1955. Studies in African linguistic classification. New Haven: Compass Publishing Company.

Heine, Bernd, and Derek Nurse, 2000. Niger-Congo, N-CL. New York: University Press of America.

Jenewari, Charles. 1989. Ijoid. In John Bendor-Samuel, pp. 105–18.

Köhler, Oswin. 1975. Geschichte und Probleme der Gliederung der Sprachen Afrikas. Die Völker Afrikas und ihre traditionellen Kulturen, ed.by H. Baumann, 137-373. Wiesbaden: Franz Steiner.

Marchese, Lynell. 1989. Kru. In John Bendor-Samuel, pp. 119–40.

Mukarovsky. 1976. A study of Western Nigritic. Vienna: Institut für Aegyptologie und Afrikanistik.

Naden, Tony. 1989. Gur. In John Bendor-Samuel, pp. 141–68.

Samarin, William. 1971. Adamawa-Eastern. Samarin, William.

1971. Adamawa-Eastern. Current trends in linguistics, Vol. VII: Linguistics in Sub-Saharan Africa, ed. by T. A. Sebeok, 213–22. The Hague, Netherlands: Mouton.

Sapir, John. 1971. West Atlantic: an inventory of the languages, their noun class systems and consonantal alternations. CTL 7. 45–112.

Schadeberg, Thilo. 1989. Kordofanian. In John Bendor-Samuel, pp. 67–80.

Stewart, John. 1989. Kwa. In John Bendor-Samuel, pp. 217–46.

Williamson, Kay. 1989a. Niger-Congo overview. In John Bendor-Samuel, pp. 3–46.

Williamson, Kay. 1989b. Benue-Congo. In John Bendor-Samuel, pp. 247–74.

Wilson, W. A. A. 1989. Atlantic. In John Bendor-Samuel, pp. 81–104.

<div align="right">SAMUEL GYASI OBENG</div>

Nigeria

The Federal Republic of Nigeria is the most populous country in Africa. This 'giant of Africa' gained independence from the British on October 1, 1960. Nigeria is located in West Africa. In the south, it is washed by the Atlantic Ocean. Its four neighbors are Nigér Republic in the north, Chad in the northeast, Cameroon in the east, and Benin Republic in the west. Presently, it has a presidential form of government. The centrally located city of Abuja is Nigeria's new capital. The Niger river and its tributary, the Benue, divide Nigeria into three parts: the northern region or Hausaland, the southwestern region or Yorubaland, and the southeastern region or I(g)boland. The names of these regions mirror the linguistic basis of their nomenclature; Hausa is the primary language of Hausaland, Yoruba is the main language of Yorubaland, and Igbo (or Ibo) is the primary language of Igboland. The importance of River Niger is evident in the name of the country, which is exemplary of a creative linguistic process, which combines the words 'Niger' and 'area'. 'Nig- area', spelled NIG-ER-IA, literally refers to the area around the River Niger.

Nigeria has the largest number of languages found in any African nation, accounting for over one quarter of Africa's languages. Nigeria's linguistic landscape is as variable as its vegetation, which ranges from swampland on the southern coast, to lush tropical rainforests, and a continuum of savannah land—guinea savannah, through sudan savannah, sahel savannah, and finally desert, studded with the mighty baobab. The flora and fauna are distinctive to the region; many plants and trees are not known to the western world; hence, they have become part of the lexicon of Nigerian English. Prime examples are *kola (nut), cam wood,* the *udala* tree, *bitter leaf* (from which *bitter leaf soup* is made), and the *acacia* tree. In *Things fall apart*, a classic bestseller that has sold over two million copies and has been translated into over 50 languages, Achebe utilizes Nigerian English, which echoes the oral tradition the text celebrates. The text is rich in proverbs, metaphor, simile, and folklore, as exemplified by the following excerpts:

a. 'Okoye said the next half a dozen sentences in proverbs. Among the Ibo, the art of conversation is regarded very highly, and proverbs are the palmoil with which words are eaten' (7).
b. 'A clan was like a lizard; if it lost its tail, it soon grew another' (171)

Stories of Igboland are recounted and significant beliefs are celebrated, such as the Week of Peace, the New Yam Festival, and the precolonial conception of time in relation to market weeks and the position of the moon.

In Nigeria, English serves multifarious functions—in administration, education, commerce, and more. The national anthem and the national pledge, the two most widely recited verses of patriotism, are both in English, and illustrate how Nigeria is conceptualized as both a fatherland and a motherland. As a result of the dynamics of English usage in different strata of Nigerian society, at least three varieties of English are discernable: the acrolect, the mesolect, and the basilect (see Bamiro 1991; Pandey 1997). Many Nigerians now use Americanisms.

Multilingualism, with its accompanying problems as regards the selection of a national language, has 'forced' English—a neutral code—to function, in many ways, as the quasinational language (Emenyonu 1989:83). English is now a permanent resident in Nigeria, a primary Nigerian language in all senses of the term. In Achebe's words, 'It has been bent and twisted to bear the burden of the African experience' (1975). It is a Nigerian brand of English (Ubahakwe 1979; Akere 1982; Bamgbose 1982; Odumuh 1987), with a distinctive local flavor. As Akere (1982) puts it,

What has happened here in Nigeria… is that the resources of a second language are superimposed on an intricate system of social and kinship relationships, and on a completely different pattern of cultural outlook and social expectations. …The English forms of address and greetings have become modified to suit local communicative needs. (97)

The educated Nigerian generally has access to the largest language repertoire and even initiates switches to the stigmatized Nigerian Pidgin (English) in conversations with peers and lesser-educated Nigerians. This is a marked difference between the 'Outer Circle' (see Kachru 1982) to which Nigeria belongs on account of the variety of English it has grown, and the 'Inner circle', where it is rather unusual to find college-educated individuals being the most multilingual and multidialectal citizens who readily and willingly switch to nonstandard varieties in exchanges with lesser-educated citizens (see Pandey 1999). Not surprisingly, code-mixing, code-switching, and speech accommodation are frequently used conversational strategies, even by Nigerian creative writers whose language use mirrors the conversational realities of Nigerian society (see Pandey 1997).

Occupying approximately three fifths of Nigeria's land area, the Niger-Benue river basin is a major source of sustenance. Its waters have also nourished and transported a variety of water-borne languages, namely, Pidgins and Creoles, and specifically Nigerian Pidgin English (NPE hereafter), the nation's linguistic lifeline or primary lingua franca. NPE still bears linguistic evidence of the earliest European contact in Nigeria. The Portuguese are known to have arrived on the coast of West Africa in the fifteenth century. Exemplary Portuguese words that constitute part an parcel of the core vocabulary of this linguistic medley (NPE) include *sabi* (comprehend, understand), *pickin* (child or children), *palava* (trouble), and *dash* (a gift or bribe, including the act of giving). The excerpt that follows, an Efik Chief's diary entry, provides some of the earliest evidence of the development of a pidginized interlanguage in Calabar: 'about 6 am in aqua Landing with small Rain morning so I walk up to see Esim and Egbo Young so I see Jimmy Henshaw come to see wee and wee tell him for go on bord …' (Forde 1956). Around Lokoja, a confluence town, different varieties of pidgin are spoken in a relatively small area. The same has been recorded for the Delta and Rivers regions, where several varieties of NPE have been documented (Marchese and Shnukal 1982; Faraclas 1985). Opinions differ regarding the conjectured expansion of NPE into a Creole in some parts of Nigeria (Donwa-Ifode 1983).

Even though English is Nigeria's official language, NPE is the preferred lingua franca; hence, it should come as no surprise that NPE is the most widely spoken pidgin in the world, with over 40 million estimated speakers, not all of whom are Nigerians (Faraclas 1996:1). Some speakers are from neighboring Cameroon and Ghana—both of which witnessed a 'brain drain' (the Nigerian English term for intellectual loss) or an exodus of immigrants to Nigeria—during Nigeria's oil boom. For the most part, these and other expatriates have acquired NPE. In the absence of a standard orthography, NPE is written the way it sounds (to the speaker). Faraclas (1996) observes that the name NPE is actually a misnomer, as it has developed trilectal varieties: a pidginized basilectal variant often referred to as 'Pidgin Propa' (i.e. 'Pidgin Proper' or PP) and frequently used by less educated or uneducated Nigerians; a creolized 'mesolect'; and a decreolized variant that is often used by educated Nigerians. Eze (1980) refers to this variety as 'Hyperanglised Pidgin' (HPP) and notes that it differs extensively from 'Pidgin Propa' in its more-Standard-English-oriented use of prepositions, as opposed to the use of 'for' to represent all prepositional functions in 'Pidgin Propa'. There are, of course, other differences, such as the more economical lexicon of 'Pidgin Proper'. Examples include:

(1) a. Im mother been come meet am <u>in</u> the office (HPP)
b. Im moda been come see am <u>for</u> (im) office (PP)

[Gloss: His/Her mother came to meet or see him/her in the office OR His/Her mother DID (indeed) come and meet him/her at the/his/her office]

(2) a. 'The thing pained me bad bad because I wanted to be big man like lawyer or doctor riding car and talking big big English' (HPP, excerpted from Saro-Wiwa's *Sozaboy*, p. 2)

[Gloss: The thing/It hurt me very badly because I wanted to be an accomplished and successful person, like a doctor or a lawyer, driving a nice car and impressing people with my command of the English language]

b. E/De ting vex me bad bad. I wan be big man like lawya or docto wey get car and tok big big English (PP)
(3) 'Everyting scatta' (from 'Zombie') [Gloss: Everything is scattered/is in disarray]
(4) 'I look the D.O. well well. Then I tell am say no be tallness go fight the war…' (from Saro-Wiwa's *Sozaboy*, p. 27)

[Gloss: I looked at/stared at the D.O. for a long time. Then I told him that when it came to fighting the war, it didn't matter how tall you were]

(5) *E get one pikin, I/Ai get four (pikin*) [Gloss: She/he has one child; I have four (children)]

(6) 'If you neva ready to carry the load, O', why put am for another person head?… Dis kain life, na so so wahala' (lyrics excerpted from 'Wait for Me' composed by King Sunny Ade and Shina Peters) [Gloss: If you aren't ready to shoulder your responsibility, why burden others with it? That approach is full of troubles]

(7) 'No money, woman no go follow you, even if you fine fine pass everybody' (from Saro-Wiwa's *Sozaboy*, p. 21)

[Gloss: If you have no money, women won't come after you, even if you are the most handsome man in the world]

(8) Wettin you de talk? (What are you talking about? OR What *thing* are you referring to?)

Nonetheless, some core features that different varieties of NPE share include an anglicized vocabulary: African grammatical and discourse structures, the nonuse of tense markers, as in (3) and (4), the zero copula and zero distinctions in gender and number (see (1b)), zero genitive forms (see (6)), reduplication—a feature of many African languages (see (4) and (7)), and multiple meanings associated with a single word (e.g. the word *chop*, which means *consume(d), eat(en), ate, digest(ed), food, a snack,* and *a bribe).* Other similarities include a mixed vocabulary that contains many words from Nigerian languages, reduced or zero articles, and demonstratives; communicative contractions such as 'tori' for 'story', 'gree' for 'agree', and 'kain' for 'kind"; communicatively expressive compounds like 'sofahead' (a creative term used to refer to one who worries too much), and coalesced creations, such as 'likam' (like him/her, likes him/her), 'wey' (who, whose, where), and 'wettin', a combination of 'what' and 'thing', as in (8).

Like Nigerian English, NPE is particularly creative, and echoes the rich Nigerian oral tradition (over 2,000 years old). Proverbs are a vital ingredient and color this code in a refreshing manner. Examples include: *Who go dash monkey banana?* (Gloss: Who gives monkeys bananas out of sheer goodwill?/There are no free lunches) and *Kaki no be leda* (Gloss: Khaki is not leather; in short, it's like comparing apples and oranges). Sayings in NPE enable their speakers to succinctly communicate a whole lot. In a culture where economy, metaphor, and wit are prized in conversation, this code is valuable. Because of its function as the language of the people, many Nigerian literary artists utilize NPE in their works (Lindfors 1974). In fact, for a long time, NPE served as a safe outlet for political expression, particularly during military regimes (Eka 1999). Exemplary works include *Sozaboy: A novel in*

rotten English written by the recently hanged Nigerian writer Ken Saro-Wiwa, who, although not a linguist, made the following sociolinguistic observation:

Both 'High Life' and *Sozaboy* are the result of my fascination with the adaptability of the English Language and of my closely observing the speech and writings of a certain segment of Nigerian society. For, as Platt, Weber, and Ho accurately observe in their book *The new Englishes* (1984), 'In some nations … the New Englishes have developed a noticeable range of different varieties. … ' Sozaboy's language is what I call 'rotten English', a mixture of Nigerian pidgin English, broken English, and occasional flashes of good, even idiomatic English. This language is disordered and disorderly. … It thrives on lawlessness, and is part of the dislocated and discordant society in which Sozaboy must live, move, and have not his being ('Author's Note').

Works by other Nigerian artists that are entirely or predominantly in NPE include 'Zombie', lyrics by Fela Kuti, one of Africa's foremost musicians, and *No longer at ease* in which the characters' use of NPE both mirrors and intensifies the neocolonial discord Achebe depicts through his artful use of varieties of Nigerian English. Some of Africa's best-known writers, playwrights, and musicians come from Nigeria, and their works have been instrumental in the development of Nigerian English. Several are world-renowned and have contributed substantially to both African literature and world literature. In 1986, Wole Soyinka became the first African to receive the Nobel Prize in literature. Chinua Achebe was awarded the Nigerian National Merit Award, Nigeria's highest award for intellectual achievement. *Arrow of God* won the New Statesman-Jock Campbell Award, and *Anthills of the Savannah* was the finalist for the 1987 Booker Prize.

Nigerian artists such as Chinua Achebe, Wole Soyinka, Ola Rotimi, Amos Tutuola, John Pepper Clark, Segun Oyewole, Tunde Fatunde, Christopher Okigbo, Gabriel Okara, and Femi Osofisan have been known to utilize NPE for literary effect—to color their characters or the plot (e.g. Amos Tutuola's *The Palm-Wine Drinkard),* to add humor (e.g. Soyinka's *Jero's Plays,* Segun Oyewole's *Katakata for Sofahead*), sarcasm, local flavor, or to give the text a linguistic facelift (see Pandey 1997). NPE is also frequently used in Onitsha Market literature, including well-known plays like Ogali Ogali's *Veronica my daughter* that has sold over 250,000 copies since 1956, and short stories written by a host of Nigerian pamphleteers (Sander 1980). Ogali uses NPE or 'uncooked English', as Veronica calls it, to portray her traditional father (Chief Jombo) and Chief Bassey, and Standard English to portray the other characters. Bomber Billy speaks a bombastic variety of English, and is ridiculed by the writer and by

one of the characters in his play, Paulina. She describes him as 'a negligible pocket radio that utters useless words'. In Sander's (1980) opinion,

> Ogali is making fun of those people in Nigerian society who, eager to demonstrate their mastery of the English language, go to any extremes in the use of polysyllabic words or the creation of new words. … Through the use of various levels of English, Ogali manipulates the response of the audience (x).

In short, NPE is frequently utilized in Nigerian literature, and its continued use contributes to its popularity. As far back as 1979, NPE was observed to have replaced Efik as the lingua franca of the Cross River area (Dunstan 1969:35).

Being a West African pidgin, NPE is of vital importance to the study of creolistics, because of its historical role in the development of anglophone pidgins and creoles. In the words of Nicholas Faraclas (1996), 'Nigerian Pidgin can be considered to be one link in a chain of English-lexifier pidgins and creoles spoken along the coast of West Africa and in African diaspora communities throughout the Atlantic Basin' (1). Advocates of the Afrogenesis hypothesis, which traces the roots of pidgins and creoles to Africa, look to West African pidgins and creoles for answers (Romaine 1994; Dalphinis 1985). According to one view, 'Created in West Africa, … work pidgins were transported across oceans where they took on new roles as lingua francas among the enslaved. …' (McWhorter 1984:240). In the case of NPE, as well as Nigerian English, many of the ingredients are African. These include the names of food (*akara balls,* fried bean cakes; *moimoi,* steamed bean cakes; *dodo,* fried plantains; *gari,* powdered cassava; *Jollof rice,* rice flavored with Nigerian peppers; *masa;* millet cakes eaten in Hausaland; *pounded yam, egusi soup,* etc.); clothing (*agbada,* the flowing robes that Yoruba men wear; *buba and wrapper*, the women's traditional outfit); local customs and festivals (*naming ceremonies* for newborns); regional celebrations (e.g. Masquerade Day, a tribute to the ancestors in the Ekiti region); and expressions of linguistic pride such as *ajibotta*, the well-known derogative Yoruba word for a butter eater or westernized Nigerian (the Nigerian English equivalent is *been-to).* Exemplary Yoruba words in NPE include *Oga* (a respectful term, roughly equivalent to 'Sir'), *katakata* (major disturbance or problem), and the honorific *O'.* Some Hausa words that have entered NPE include *wahala* (trouble) and *haba* (Goodness! or Listen up!). Igbo words in NPE include *sef* (a discourse marker), *yanga* (boasting or boastfulness), and *na* (a particle with multiple meanings, such as 'it is', and the attention-getter *Na wa, O'!*, a popular expression of shock).

The vast majority of Nigeria's languages—including Yoruba and Igbo, two of the three major languages—belong to the Niger-Congo family. Estimates vary regarding the exact number of indigenous languages thriving in Nigeria. While Bamgbose (1991:2) and McArthur (1992) put the number at 400, Katzner (1995) estimates that there are 'about 250' (352). While we cannot rule out the threat of language endangerment arising from the central place that English occupies in Nigeria's linguistic horizon, the discrepancy in estimates has a lot to do with differences in opinions regarding what constitutes a language and what constitutes a dialect. For instance, Ibibio, Efik, and Annang, being mutually intelligible, are viewed by many Nigerians as one and the same language (Dunstan 1969:35), namely Efik/Calabar (or Ibibio to some Nigerians), which has acquired the status of a Standard dialect, on account of its use in broadcasts and textbooks. Some Nigerians do, however, regard them as separate languages. According to some, Ikwer(r)e/Ikwerri has 12 dialects. Many non-Ibibio speakers regard Qua, Oron, Eket, Okobo, and Ibuno as dialects of Ibibio, while others view them as lesser-used languages. While some view Kolokuma and Ijo as dialects of the same language, others might not, as Nigerian codes of communication have distinct names. Most have several names. Isoko, for instance, is also called Biotu, Sobo, and Igaba; yet, some consider the last two names to be offensive.

To most Nigerians, dialects tend to coincide with geographical boundaries or specific urban locations. Chief dialects identified for Eksako, for instance, are said to be those of Auchi, Aviele, Ekperi, South Ibie, Uzairue, and Weppe Wano—names of towns and cities, and of specific 'clans' (Dunstan 1969:47). The three main dialects of Fula/Fulani found in Nigeria also coincide with the names of major cities or geographical terrains, namely *Sokoto, Adamawa,* and the *central dialect* of Kano, Bauchi, Gombe, Katagum Division, and Jos (Dunstan 1969:57).

Most educated Nigerians speak at least three languages, English (the official language), one or more of the major regional languages, a local language, and NPE, not to mention the other dialect(s) they might switch to in the middle of a conversation. For most Qua speakers, for instance, Efik is a second language, Igbo the third, NPE the fourth, and (Nigerian) English the fifth.

In Nigeria, ethnicity and linguistic identities are inextricably intertwined. McArthur (1992) recognizes eight major ethnic groups, namely 21% Hausa, 21% Yoruba, 18% Igbo, 10% Fulani, 6% Tiv, 5% Kanuri, 5% Ibibio, 4% Edo, and 10% minorities. Because of the close ties between the Hausa and the Fulani, it is not unusual to find these groups conjoined under one

label, namely, the Hausa-Fulani. The minority ethnic groups would include, among others, the Ogoni, Isoko, Izon, Kwale, Degema, Itsekiri, Ibuno, Izi, Ewe, Andoni, Amo, Angas, Birom, Chip, Chella, Efik, Mada, Nabor, Nembe, Tera, Yergam, Gwari, and Gonja peoples of Nigeria. The names of their languages tend to coincide with their ethnic descriptors, further emphasizing the linguistic basis of ethnic identity in Nigeria. Each ethnic group has its own unique customs and language, making for a diverse group of Nigerians; yet, the minor languages tend not to be as carefully studied. Some (Fula, Tiv, Igala, Efik, and Edo) enjoy a more functional status than others, because of their use in the media. The nation's linguistic and cultural diversity has a lot to do with its historical location at the heart of transcontinental migration routes. The greatest concentration of languages and ethnic groups is found in Central Nigeria. One has only to travel to Kwara State or Illorin to get a feel for the linguistic diversity.

The most widely spoken indigenous language is Hausa (Katzner 1985; McArthur 1992). This is followed closely by Yoruba (18% of the population or some 20 million speakers), then Igbo (11% or some 15 million speakers), Fulani (approximately 8 million speakers), Kanuri (3 million), Efik and Ibibio (3 million), Tiv (spoken around Gboko by some 2 million), Ijo or Ijaw (2 million), Edo (one million), Urhobo (half a million), Nupe (in the Gur subfamily of Niger-Congo languages, with roughly 500,000 speakers), Idoma (some 250,000 speakers), and Eksako (also spelled Esako and Etsako(r), an Edo language with some 120,000 speakers). These numbers are more estimates than up-to-date survey findings of 'native speakers' of these languages, as representative language surveys are hard to administer in this complex multilingual environment, and such surveys have rarely been administered (see Bamgbose 1995:34).

Hausa, a Chadic language that contains many words borrowed from Arabic, is spoken by roughly 27% of the population or some 40 million Nigerians. It is the primary language or mother tongue of at least 25 million Nigerians (Katzner 1985:288). The Kano dialect is regarded as the Standard variety. The number of Hausa speakers has clearly grown; in 1979, an estimated 12 million Nigerians were native Hausa speakers (Dunstan 1969:73). Hausa is the lingua franca for the bulk of northern Nigeria and also plays the role of a regional language in West Africa. It is, in fact, the most widely spoken African language in West Africa, with speakers found in neighboring countries, including Nigér (as many as five million speakers), Ghana, Togo, Benin, Mali, and Senegal. The Hausa Language Board, established in 1955, has assisted in the standardization of the language. Prior to independence, Hausa enjoyed official language status alongside English in what was then the Northern Region. At that time, Hausa was written in the Arabic script known as Ajami, but now the Roman script is used.

Yoruba, a paradigm Kwa language and a subgroup of the Niger-Congo family, is spoken in southwestern Nigerian, from Illorin (Kwara State) down to Akure, Ado-Ekiti, and Igede-Ekiti (Ekiti State), Ibadan (Oyo State), and Lagos. It is Nigeria's second major language. This language and the rich mythology (of Sango and Ogun) constitute seminal ingredients in many of Wole Soyinka's works. The mythology is also resonant in the Afro-Caribbean beliefs transported to the Caribbean (see Burnett 1986). Some 20 dialects of Yoruba can be heard in Nigeria, and one, 'based largely on the Oyo dialect' (Dunstan 1969:80), is the Standard. Yoruba is also spoken in some other West African nations, including parts of Ghana, Togo, and Benin. Tonality is mirrored in its orthography. Three letters and their corresponding sounds are noteworthy: o, pronounced as in 'sought' (e.g. oba); é pronouced as in 'let' (e.g. *éjé* which means blood); and s, pronounced /sh/ as in 'ekuse' (i.e. Well done!). In orthography, the acute and grave accents indicate tone (not stress).

Igbo, another Kwa language like Yoruba and Ewe, is Nigeria's third major language. It is spoken in southeastern Nigeria, in Port Harcourt, Enugu, Calabar, Onitsha, Owerri, and surrounding areas. The three main dialects that are recognized are the Central, Owerri, and Umuahia dialects. Prior to 1961, its orthography was disputed (Dunstan 1969:85). The official orthography, adopted by the Onwu Committtee, is now compulsory in the School Certificate Igbo examination. Igbo features regularly in the works of Chinua Achebe. In *Things fall apart*, some 37 Ibo words and phrases are woven into the fabric of the text. Many have no translation equivalents, including *chi* (personal god), *ogbanje* (which Achebe describes as a 'changeling' that will keep dying until its 'iyi-uwa' is first dug up and destroyed), *uri* (a betrothal day, after the bride-price has been paid), *Nne* (mother, including one's nonbiological mothers in a polygamous household), *umunna* (kinsmen), and *ndichie* (one's elders). These indigenous words bring the richness of precolonial Igbo society to life. From a sociolinguistic standpoint, they constitute prime examples of the Sapir–Whorfian hypothesis, which illustrates how specific cultural concepts are embodied in words.

Fula(ni) is generally termed Fulfulde in Nigeria. This Niger-Congo language (the West Atlantic branch) of the Fulani/Fulbe people is of great historical significance, as the Fulani are 'a people of great antiquity' (Katzner 1995:289) and have heavily influenced the sociolinguistics of northern Nigeria, where over eight

million of the estimated 15 million Fulani people reside. Many are cattle-herders, and follow the nomadic and seminomadic lifestyle they are accustomed to. Areas in which Fulani is spoken in Nigeria include Sokoto, Kano, Kaduna, Plateau, Bauchi, and Kwara States. In Mauritania, Gambia, Benin, Ghana, Mali, Burkina Faso, Chad, Nigér, and Cameroon, where it is also spoken, it is called Fula. In Senegal, it is called by its other name, Pulaar, and in Guinea-Bissau, it is termed Pulle. The root morpheme in all these names is {Ful-} or {Pul-} from which the Germanic name Fula(ni) and the Frenchified 'Peul' originated. Its regional spread makes it a frequently used lingua franca in West Africa.

The national language question has received quite a bit of attention since the mid-1970s. In *Language and the nation,* Ayo Bamgbose observes that:

> The language question in Sub-Saharan Africa arises from the fact that not only are most of the countries multilingual, the colonial experience has led to the importation of foreign official languages which have taken on the roles of national communication, unification, administration and medium of education from early or late primary to university level. Thus, the existing multilingual situation is compounded by the addition of imported languages whose strength does not lie in numbers of speakers, but rather in the superior roles assigned to them (1).

During the colonial period, western education was restricted to only a small segment of the population that played a role in the 'indirect rule' in place; hence, even today, only the elite uses Standard English. British administrative policy in Nigeria relied on empowered Nigerians—primarily Chiefs and Emirs—to serve as intermediary rulers. English usage and English-medium western education were most widespread in parts of the country where Christian missions were established (Ajayi and Crowder 1985:68). Islam (roughly 47% of the population) and Christianity are the two primary religions, with an estimated 19% of the populace being 'traditionalists' (McArthur 1992). As Christianity was resisted in Islamic northern Nigeria, where the Jihad or Holy War led by Uthman dan Fodio in the early nineteenth century had left its religious mark, the discrepancy in English usage between northern and southern Nigeria has persisted to this day. So pronounced were the differences on the eve of independence that there was even a fear that those with an English-based Western education 'would dominate the economic and political life' (Ajayi and Crowder 1985:68) of the new nation. The establishment of major regional universities can be attributed, in part, to this regional-linguistic competition. As the University of Ibadan, Nigeria's oldest institution, was garnering a lot of (inter) national attention, many Nigerians saw the need for a university in Igboland and another in Hausaland. At the wake of

independence, then, several universities were established in Nigeria, most notable among them The University of Nigeria in Nssuka, eastern Nigeria, Ahmadu Bello University, located at the heart of Hausaland, and the University of Ile-Ife, located in Ife, a very important city in Yorubaland, and home of the Oba of Ife, one of the most highly regarded Chiefs in the Yoruba chieftaincy. Ahmadu Bello University (better known as ABU) was named after (Sir) Ahmadu Bello who was northern Nigeria's first Premier. A resident of Hausaland, and a well-known political figure, northerners (called 'Gambaris' by some southerners) named the University after him, as a tribute to his accomplishments. The University is located in Zaria, a famous northern city that played a pivotal role in the intricate emirate system that was in place prior to the colonial period.

English and NPE usage declines steadily as one moves north of the town of Illorin. Instead, one tends to hear more Hausa and Fulani in everyday conversations and even in service-counter exchanges. Koranic schools and after-school Arabic classes are commonplace in the north, where Arabic is a language of great (religious) importance, although rarely used in conversations. In contrast, in the open markets in the Yoruba-speaking Ekiti area or in Onitsha, which is famous for its market, NPE is the preferred transactional code, ideal for bargaining. So important is NPE in the markets of Yorubaland and Igboland that mastery of NPE gives buyers a competitive advantage over non-NPE speakers, as haggling is a culturally acceptable and expected discourse practice. One who does not speak the language of the seller is usually at a disadvantage. Some vendors can be heard speaking what is called 'Broken English' in Nigeria, and mainly in their limited-but-meaningful exchanges (e.g. 'Oga, buy tomato! Good Good!') with the Oyĩbo(s), the Yoruba term used to identify outsiders or non-Nigerians. Some traders and children use the compound term 'Oyĩbo man'. This term literally means 'white man' or 'white person', but its use is not limited to Europeans alone. Also, it is not intended as an insult, particularly when it is followed by the respect marker 'O(h)', although it could easily be misconstrued as such, particularly when one is surrounded by a crowd chorusing 'Oyĩbo O''. The term should be interpreted as more of a congregational announcement of the presence of a visitor whom one would like to honor by acknowledging. 'O'' is an honorific term in Yoruba, which means roughly 'respectful Sir or Ma'am', and has been adopted by NPE. The equivalent term for expatriates in Hausa is 'Baturé' (masculine) or 'Baturiya' (feminine).

It would be a mistake to view Nigeria as an ESL environment, because it is an English-as-an-indigenized-language context (Bamgbose 1995). Even

though English is the medium of instruction from third grade onward, most Nigerians speak Nigerian English, which differs from Standard British or American English. In the past 10 years, a matter of grave concern to Nigerian educators has been the 'mass failure syndrome' (Bamgbose 1995:130–52) or the increasing number of failing grades in English obtained in the main college-entry examinations, namely, the WAEC or the West African School Certificate Examination, and the JAMB or the Joint Admissions and Matriculation Board Examination. These examinations are patterned along the colonial, British variety, yet the students hear and learn a Nigerian variety. Many of the idioms that students are tested on are, in fact, foreign to the culture and environment. Examples include the following italicized items: a finger in every *pie,* comparing *apples* and oranges, looking for a needle in a *haystack,* over the *grape vine,* and *penny* wise, *pound* foolish. Students are often tested on noncount nouns like *information,* even though most Nigerians pluralize these. The end result is a mismatch between the variety of English that is learned, spoken, heard, and taught, on the one hand, and what is tested. Admission into college is contingent upon a credit pass in English, and yet a curricular change is unlikely, as administrators do not recognize Nigerian English.

In a report submitted to the Federal Ministry of Information, the Nigerian Public Service Review Commission drew attention to the need for a national language:

> An overriding problem, which affects the public service as it does all aspects of society, is that of language. Nigeria shares with many developing, and some developed nations, the lack of an indigenous *lingua franca.* What this means for efficiency in the conduct of government business is rarely even thought about perhaps because there seems to be no immediate answer. But it is perfectly clear to the careful observer that below the top-most levels in the various sectors of society, people are conducting their business in a language which, in varying degrees, they have not in fact mastered (6).

Since that time, some of the major motions that have been tabled include the idea of a trilectal national language policy, a rotational language policy (which also draws on the three regional languages, one by one), and the creation of a single conglomerate language, like Esperanto, which would be called WAZOBIA, and which would combine elements of Hausa, Yoruba, and Igbo. Another idea that has received considerable attention is the pan-Africanist idea of adopting the most widely spoken African language, namely, Swahili. This proposal has been led by Wole Soyinka. In the last few years, French has been added to the list of contested languages. In general, Nigerians in favor

of a national language have advanced the idea of an indigenous national language.

Special programs have been designed to reduce tribalism and foster national unity. One such measure is the Nigerian Youth Corp Service, better known as the NYSC, which is a year-long commitment required of all graduates. It requires them to serve in a region that is linguistically different from their own, so that they learn to be more tolerant of other ethnic and linguistic groups. This year of service almost always ensures a process of assimilation. Most 'youth corpers' end up acquiring yet another language—usually a major regional language—during their NYSC year. The prop on which this program rests is essentially sociolinguistic. Another multilingual measure is the attempt to revive the local languages, particularly the three major languages that constitute the cornerstones of Nigeria's linguistic 'stool' (a symbol of the seat of government in the precolonial Chieftaincy era). This includes making them instructionally viable (Rufa'i 1991; Arohunmolase 1998, 1999) through the development of textbooks, teacher preparation, and so on. The WAEC now offers certification examinations in some Nigerian languages.

References

Achebe, Chinua. 1958. Things fall apart. London: Heinemann.

Achebe, Chinua. 1964. Arrow of God. London: Heinemann.

Achebe, Chinua. 1987. Anthills of the Savannah. London: Heinemann; 2nd edition, 1989.

Achebe, Chinua.1975. Morning yet on creation day. New York: Doubleday Press.

Ajayi, Ade, and Michael Crowder (general eds.) (E. Dunstan, linguistics ed.) 1985. Historical atlas of Africa. Cambridge: Cambridge University Press.

Akere, Funso. 1982. Sociocultural constraints and the emergence of Nigerian English. New Englishes, ed. by J.B. Pride, 85–9. Rowley, MA: Newbury House.

Altback, Phillip, G. (eds.) 1999. Publishing in African languages: challenges and prospects. Asmara; Trenton, NJ: Africa World Press, Inc.

Amkpa, Awam. 1998. Nigerian dramatic literature: the examples of Femi Osofisan and Tunde Fatunde. The growth of African literature: 25 years after Dakar and Fourah Bay, ed. by Edris Makward, et al. Asmara: Africa World Press, Inc.

Arohunmolase, Oyewole (ed.) 1998. Nigerian languages for national development and unity. Trenton, NJ: Africa World Press, Inc.

Avery-Coger, and Greta M.K. McCormick (eds.) 1988. Index of subjects, proverbs, and themes in the writing of Wole Soyinka. New York: Greenwood Press.

Bamgbose, Ayo. 1982. Standard Nigerian English: issues of identification, In Kachru, pp. 148–61.

Bamgbose, Ayo. 1991. Language and the nation: the language question in sub-Saharan Africa. Edinburgh: Edinburgh University Press.

Bamgbose, Ayo, et al. (eds.) 1995. New Englishes: a West African perspective. Ibadan: Mosuro.

Burnett, Paula. 1986. The Penguin book of Caribbean verse in English. Harmondsworth: Penguin Publishers.

Dalphinis, Morgan. 1985. Caribbean and African languages: social history, language, literature, and education. London: Karia Press.

Donwa-Ifode, S. 1983. Is Nigerian Pidgin English Creolizing? Journal of the Linguistic Society of Nigeria 2. 199–203.

Dunstan, Elizabeth (ed.) 1969. Twelve Nigerian languages. New York: Africana Publishing Corporation.

Eka, David. 2000. Aspects of language in Ken Saro-Wiwa's Sozaboy: a novel in rotten English. Before I am hanged, ed. by O. Okome. Trenton, NJ: Africa World Press, Inc.

Emenyonu, Ernest. 1989. National language policy in Nigeria: implications for English teaching. Language planning and English language teaching, ed. by C. Kennedy, 82–91. Oxford: Oxford University Press.

Eze,Smart, N. 1980. Nigerian Pidgin sentence complexity, Band 8. Wien, Germany: Afro-Publishers.

Faraclas, Nicholas. 1985. Rivers Pidgin English: tone, stress, or pitch-accent language? Language in African culture and society. Studies in the Linguistic Sciences 14(2), ed. by E. Bokamba. 67–76.

———. 1996. Nigerian Pidgin. New York: Routledge.

Fatunde, Tunde. 1985. No food, no country. Benin City, Nigeria: Adena Publishers.

Forde, C. D. 1956. Efik traders of old Calabar. London: International African Institute.

Kachru, Braj (ed.) 1982. The other tongue: English across cultures. Urbana: University of Illinois Press; 2nd edition, 1992.

Katzner, Kenneth. 1995. The languages of the World. New York: Routledge.

Lindfors, Bernth (ed.) 1974. Dem-say: interviews with eight Nigerian writers. Austin, TX: University of Texas at Austin Press.

———. 1997. African textualities: texts, pre-texts, and contexts of African literature. Trenton, NJ: Africa World Press.

Maja-Pearce, Adewale (ed.) 1994. Wole Soyinka: an appraisal. Ibadan: Heinemann.

Marchese, L., and A. Shnukal. 1982. Nigerian Pidgin English of Warri. Journal of the Linguistic Society of Nigeria 1. 213–19.

McWhorter, John, H. 1984. The missing Spanish Creoles: recovering the birth of plantation contact. Berkeley: University of California Press; 2nd edition, 2000.

Odumuh, E. 1987. Nigerian English. Zaria: Ahmadu Bello University Press.

Osofisan, Femi. 1975. A Restless run of locusts. Ibadan: Onibonoje Press.

Oyekunle, Segun. 1983. Katakata for Sofahead. London: Macmillan Education.

Pandey, Anita. 1997. The Pragmatics of Code Alteration in Nigerian English. Studies in the Linguistic Sciences 25(1). 75–117.

Pandey, Anita. 1999. Code alteration and Englishization across cultures: cyclic differences. Paper invited for presentation at the International Association for World Englishes (IAWE) Convention, Singapore, and invited for inclusion in The Three Circles of English by the Conference Organizer, Edwin Thumboo, University of Singapore.

Romaine, Suzanne. 1994. Language in society: an introduction to sociolinguistics. Cambridge: Cambridge University Press.

Rufa'i, Abba (ed.) 1991. Twentieth Year Commemorative Symposium of the Center for the Study of Nigerian Languages (1989: Bayero University), Nigerian Languages Yesterday, Today, and tomorrow: Proceedings of the Symposium. Kano, Nigeria: Bayero University Press.

Sander, Reinhard. 1980. Veronica my daughter and other Onitsha plays and stories by Ogali Ogali. Port Harcourt: Three Continents Press, Inc.

Saro-Wiwa, Ken. 1985. Sozaboy: a novel in rotten English. Port Harcourt: Saros Publishers.

Ubahakwe, Ebo (ed.) 1979. Varieties and functions of English in Nigeria. Ibadan: African Universities Press.

ANITA PANDEY

See also **Bilingualism; Code-Switching; Hausa and Chadic Languages; Igbo and Igboid Languages; Niger-Congo; Official Language Selection; Pidgins and Creoles; Sapir–Whorf Hypothesis; Yoruba and Yoruboid Languages**

Nilo-Saharan Languages

The idea of a Nilo-Saharan (NiSa) phylum among African languages alongside Afro-Asiatic, Niger-Congo, and Khoisan was brought into existence by Joseph H. Greenberg (1963). On the basis of 161 lexical and 29 grammatical sound–meaning similarities, he suggested genetic unity among 82 languages, which had been treated as 22 separate language units before by Tucker and Bryan (1956, 1966) and for some of which genetic unity had never been suggested (Table 1).

On the level of language units as presented in Tucker and Bryan (1956), Greenberg proposed some significant changes, most of which have been accepted since and partly also substantiated in detailed studies:

(a) The inclusion of Mimi into the Maban group.
(b) The combination of Tucker and Bryan's 'Nilotic' and 'Nilo-Hamitic' into one group with three branches: Western, Eastern, and Southern Nilotic.
(c) The combination of Moru-Mangbetu and Bongo-Bagirmi into a Central Sudanic group.

Greenberg's fusion of Gule and Koma into one group, however, has been revised. Bender (1996) suggests again two units although differently subdivided, i.e.

TABLE 1 Language Groups of Tucker and Bryan (1956) and Greenberg (1963)

Tucker and Bryan (1956)	Greenberg (1963)		
	1. <u>Songhai</u>		
2. Eastern Sudanic	2. <u>Saharan</u>: (a) Kanuri, Kanembu; (b) Teda, Daza; (c) Zaghawa [today also: Beria], Berti		
10. Mimi 11. Maba group	3. <u>Maban</u>: Maba, Runga, Mimi (of Nachtigal), Mime (of Gaudefroy-Demombynes)		
13. Fur	4. <u>Fur</u>		
20. Nubian group			1. Nubian group
27. Didinga-Murle group			2. Murle, Logarim, etc. [today: Surmic]
21.Barea	5. <u>Chari-Nile</u>		3. Barea [today: Nara ~ Nera]
24. Tabi		A.	4. Ingassana (Tabi) [today also: Gaam/Jebel]
15. Nyimang group		Eastern	5. Nyima, Afitti
16. Temein group		Sudanic	6. Temein, Teis-um-Danab
12. Tama group			7. Merarit, Tama, Sungor
14. Daju group			8. Daju group
30. Nilotic			9. Nilotic
31. Nilo-Hamitic			
29. Teuso			10. Nyangiya, Teuso [today: Kuliak]
1. Moru-Mangbetu		B. Central Sudanic	
2. Bongo-Bagirimi			
23. Berta		C. Berta	
22. Kunama		D. Kunama	
25. 'Gule' [extinct] 26. Koma group	6. <u>Coman</u>: Koma, Ganza, Uduk, Gule, Gumuz, Mao		

I. Koman: T'wampa (Uduk), Gule, Komo, Kwama and Opo (Shita) and

J. Gumuz (dialect cluster).

Several morphological sound–meaning similarities used by Greenberg (1963) to define Nilo-Saharan are still major arguments regarding Nilo-Saharan unity, among them:

—independent subject pronouns, especially in the singular, show the vowels *a* (1st person), *i* or *u* (2nd person) and *e* (3rd person),

—*ma* and *ko* as relative and adjective formant,

—*n/k* singular–plural alternation,

—*t/k* singular–plural alternation,

—nominal derivational prefix *k-* ('movable' *k-*),

—verbal dative affix *-kV-*,

—causative in *t*

Several morphological sound–meaning correspondences proposed by Greenberg to support Nilo-Saharan unity, however, are also found in neighboring Afro-Asiatic (e.g. negation in *b* or *m*) or in Niger-Congo languages, thus weakening the argument and pointing to the general difficulty of separating areal from genetic features.

Most recently, only two comprehensive studies on the genetic classification of Nilo-Saharan since Greenberg's have been published, i.e. Bender (1996) and Ehret (2001). The comparison of the results presented in these studies highlight significant features of the current state of discussion in the field of Nilo-Saharan genetic studies:

—The idea of a Nilo-Saharan phylum remains widely accepted, although the external boundaries of Nilo-Saharan as well as significant parts of its internal subdivision remain highly controversial.

—Greenberg's proposal of a Chari-Nile group has been rejected as a valid genetic unit, while the genetic unity of Eastern Sudanic and of Central Sudanic remains unchallenged.

—The lower-level units (Tucker and Bryan's language units as revised by Greenberg) are also largely uncontroversial.

—The most substantial progress since Greenberg has been made regarding language documentation, genetic classification, and reconstruction within these lower-level units.

External boundaries

Discussions regarding the external boundaries of Nilo-Saharan relate particularly to the following topics:

Songhay cluster: Its geographic distance from the other Nilo-Saharan languages and long-standing contact with Mande (Niger-Congo) and Berber (Afroasiatic) languages continue to cast doubt on the Nilo-Saharan affiliation suggested by Greenberg (1963) and supported

by Bender (1996) and Ehret (2001). Nicolai (1990) reconstructs it as a post-Creole with a Berber base, while some other scholars follow an old argument by Delafosse and discuss Mande affiliation.

Kadugli-Krongo (Kadu): This group appears in Greenberg (1963) under the name of Tumtum as one of five Kordofanian branches despite divergent properties. Following suggestions made by Dimmendaal (1987) among others and supported by the fact that this group shows several of the typical Nilo-Saharan features (1sg pronoun *a'a*, 2sg pronoun *ʊ'ʊ*, movable *k*, *n/k* plural formation, etc.), Bender (1996) includes Kadugli-Krongo in his Core of Nilo-Saharan, while Ehret (2001) rejects any Nilo-Saharan affiliation.

Shabo (formerly: Mikeyir): The genetic position of this rudimentarily documented language of SW Ethiopia remains unclear. Bender proposes a Nilo-Saharan affiliation, which Ehret rejects.

Meroitic: The extinct language of the ancient Meroitic Empire (Sudan) preserved by a number of written records has been linked to various genetic groups and phyla, among them Afro-Asiatic, Tokharian, Saharan, and Eastern Sudanic. This latter affiliation was also supported by Greenberg, and has most recently again been substantiated by Claude Rilly, among others.

Relation to Niger-Congo: Due to a number of lexical and morphological similarities between Nilo-Saharan and Niger-Congo languages, Gregersen (1972) proposed to combine these two phyla into a single one, i.e. Kongo-Saharan. His line of argument has been taken up by Roger Blench in recent years, who attempts to establish Niger-Congo as a branch of Nilo-Saharan.

Internal subgrouping

General disagreement also characterizes higher-level units within Nilo-Saharan. Greenberg's subgrouping (Table 1) has only partly been accepted. The reason for this derives not so much from the still highly fragmentary documentation of numerous Nilo-Saharan languages, but from their generally great internal lexical and grammatical heterogeneity. This feature may point to a history of thousands of years for the whole phylum. Detailed studies of lower-level language units, however, also reveal continuous episodes of heavy language contact, thus complicating lexical and morphological reconstruction and hence higher-level grouping considerably.

Based on morphological as well as on lexical comparison, Bender (1996) proposes four branches of Nilo-Saharan: (A) Songhay, (B) Saharan, (K) Kuliak, called 'Outliers', and the rest, called 'Satellite–Core', which he further subdivides as indicated in Table 2.

Ehret's (2001:88f, 70f) alternative genetic classification of Nilo-Saharan uses phonological and lexical isoglosses derived from an extensive comparative lexical database, but also considers grammatical and derivational morphemes as far as they are accessible through language descriptions. His result is presented in a mainly bilaterally branching tree, with one specific small language group or language branching off at each level and set against the rest, which receives novel geography-based labels. In the following condensed representation of his result, previously known group and language labels are underlined (see graph).

The extreme lexical heterogeneity of Nilo-Saharan languages, even in the most basic vocabulary, is paralleled by a large grammatical diversity, and hardly any of these grammatical features is specific to the phylum.

Nilo-Saharan languages are tone languages with two to four underlying level tones, which are used for grammatical as well as lexical purposes. The number of tones and their main functional domain correlate to a considerable degree with the morphological type of the respective language. Five-vowel systems with

TABLE 2   Bender's (1996) Genetic Classification of Nilo-Saharan

Nilo-Saharan
- Koman
- Sudanic
  - Central Sudanic
  - Northern Sudanic
    - Kunama
    - Saharo-Sahelian
      - Saharan
      - Sahelian
        - Songhay
        - For[Fur]
        - Eastern Sahelian
          - Maban
          - Eastern Sudanic
            - Astaboran
              - Nara
              - W-Astaboran
                - Nubian
                - Tama
            - Kir-Abbaian
              - Jebel
                - W-Jebel
                - Bertha
              - Kir
                - Nuba-M.
                  - Temein
                  - Nyimang
                - Daju
                - Surma-Nilotic
                  - Surmic
                  - Nilotic
            - Kuliak
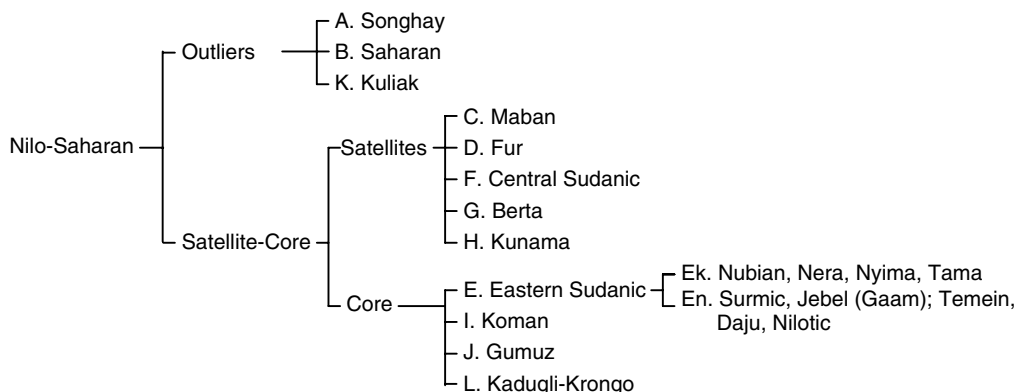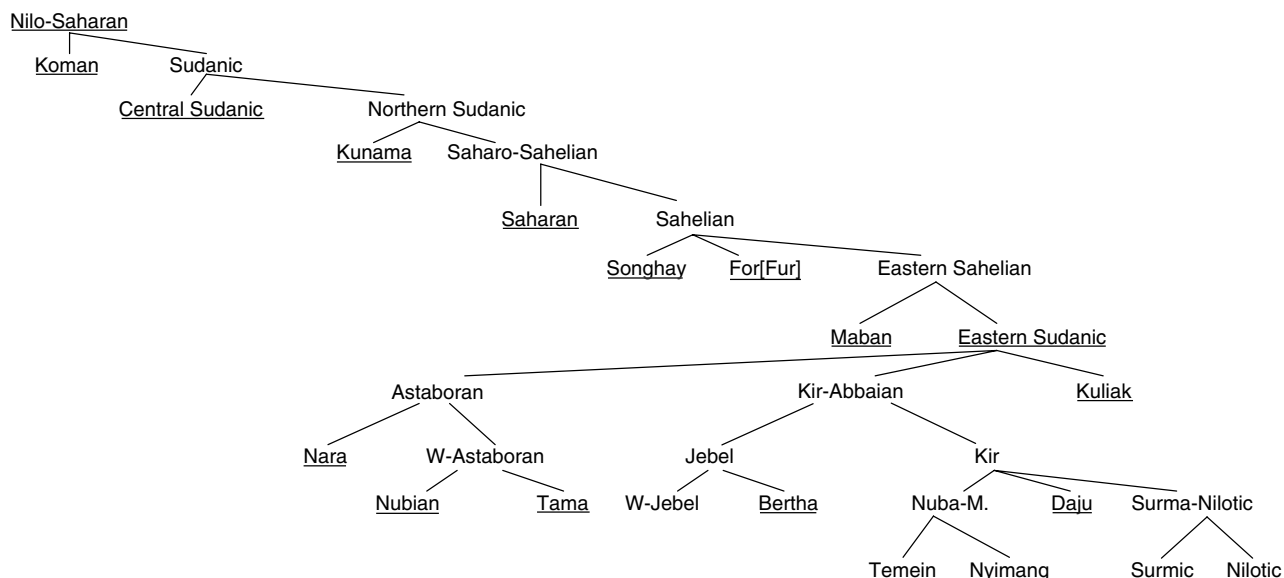
Graph. Ehret's (2001) genetic classification of Nilo-Saharan.

advanced tongue root (ATR) vowel harmony, often with contrastive length, are frequently synchronically or diachronically reconstructable. The consonant systems show either four or five (e.g. Gaam, Berta, Koma, Western Nilotic) places of articulation, which partly also extends to the nasal. Ehret (2001:16) claims five places of articulation for Proto-Nilo-Saharan in the case of plosives and four in the case of nasals. He also suggests a complex set of plosives for Proto-Nilo-Saharan, with voiced implosives and explosives, voiceless aspirated and unaspirated as well as glottalized plosives. Although all of these sets do occur in Nilo-Saharan languages synchronically, only Twampa (Uduk) comes close to such a system.

The morphological type of Nilo-Saharan languages ranges from largely isolating (e.g. Central Sudanic languages), to agglutinative (e.g. Southern Nilotic) and highly fusional (e.g. Kanuri, partly Western Nilotic). Languages of the same genetic unit often belong to different morphological types, which possibly has to be interpreted as an outcome of substantial language contact. Nonisolating Nilo-Saharan languages are mostly characterized by complex verbal derivational morphology, while isolating Nilo-Saharan languages tend to make use of serial verb construction for derivational and morphosyntactic purposes. Nominal morphology is equally heterogeneous, with some languages having gender (e.g. Eastern Nilotic), others not. A morphologically tripartite number system, in which either a singular/singulative or a plural form or both are morphologically derived from a nominal root, is widespread in Nilo-Saharan and may even be a potential isogloss.

TABLE 3 Nilo-Saharan Languages: Languages and Speakers Per Country

Country	No. of NiSa Languages	No. of NiSa Speakers	% of Population
Kenya	~ 6	7.5 Mio.	**29.2**
Sudan	**~ 64**	7.2 Mio.	**29**
Uganda	**~ 20**	5.6 Mio.	**25.5**
Nigeria	6	3.5 Mio.	3.5
DR Congo	**22**	3.0 Mio.	7
Chad	**~ 33**	2.6 Mio	**40.6**
Niger	~ 7	2.7 Mio.	**32.5**
Tanzania	~ 4	858 000	0.3
Mali	1	640 000	6
Ethiopia	**16**	376 000	0.7
CAR	13	223 000	6.6
Eritrea	2	210 000	5.7
Cameroon	2	150 000	1.2
Burkina Faso	2	125 000	0.8
Egypt	1	100 000	0.1
Benin	2	50 000	0.9
Libya	2	9 000	0.1

Syntactically, all the common basic word orders with its concomitant features occur, namely SOV (e.g. Kanuri, Nubian), VSO (e.g. East and South Nilotic, Kuliak, partly Kadugli-Krongo), and SVO. Morphological and partly also syntactic ergativity has been reported for Jur-Luwo, Anywa, Päri, Shilluk (Western Nilotic), and Toposa (Eastern Nilotic).

Demography

Numbers given for Nilo-Saharan languages range from 80–90 (lumpers) up to 199 (splitters). The area in which they are spoken stretches from the river Nile in the

TABLE 4 Numerically Dominant Nilo-Saharan Languages

Speakers	Language	Sub-division	Country
3–4 Mio.	Songhai	Songhay	Niger, Mali, Nigeria, Burkina, Faso, Benin
	Kanuri	Saharan	Nigeria, Niger, Cameroon, Chad, Sudan
	Dholuo	Western Nilotic	Kenya, Tanzania
	Dinka	Western Nilotic	Sudan
> 2 Mio.	Lwo (Acholi + Lango)	Western Nilotic	Uganda, Sudan
	Kalenjin	Southern Nilotic	Kenya
> 1 Mio.	Lugbara	Central Sudanic	DR Congo, Uganda
	Alur	Western Nilotic	DR Congo, Uganda
	Teso	Eastern Nilotic	Uganda, Kenya
	Maa	Eastern Nilotic	Kenya, Tanzania
	Nuer	Western Nilotic	Sudan, Ethiopia
> 500,000	Ngambay	Central Sudanic	Chad, Cameroon, Nigeria
	Lendu	Badha/Lendu	DR Congo, Uganda
	Mangbetu	Mangbetu-Balese	DR Congo
	Fur	Fur	Sudan, Chad
> 250,000	Maba	Maban	Chad
	Teda-Daza	Saharan	Chad, Niger, Libyen
	Masalit	Maban	Chad, Sudan
	Bari	Eastern Nilotic	Sudan, Uganda, DR Congo
	Karimojong	Eastern Nilotic	Uganda
	Nile Nubian (Nobiin)	Nubian	Sudan, Egypt

northeast up to Lake Chad and—in the case of Kanuri and Songhay—further west to northeastern Nigeria and to the Mali-Niger-Burkina Faso border area, respectively. Nilo-Saharan languages are spoken as a first language in 17 African countries by over 30 Mio. speakers (Table 3).

The highest number of speakers are found in Kenya (7.5 Mio.), Sudan (7.2 Mio.), Uganda (5.6 Mio.), Nigeria (3.5 Mio.), DR Kongo (3.0 Mio.), Niger (2.7 Mio.), Chad (2.6 Mio.), but these speakers represent significant portions of the population only in four countries, namely Chad (41%), Niger (32%), Sudan (29%), Kenya (29%), and Uganda (25%). With respect to the number of Nilo-Saharan languages per country, Sudan comes first (~64), followed by Chad, DR Congo, Uganda and—despite the overall low number of NiSa speakers—Ethiopia (16) and Central African Republic (13).

Table 4 presents major Nilo-Saharan languages in terms of number of speakers and geographic distribution. As indicated, most of them are cross-border languages, thus making language-planning activities comparatively difficult.

References

Bender, L.M. 1996. The Nilo-Saharan languages—a comparative essay. (LINCOM Handbooks in linguistics, Vol. 6.) München & Newcastle: Lincom Europe.

Dimmendaal, Gerrit J. 1987. Krongo: between universal, areal and genetic norms. Journal of African Languages and Linguistics 9. 161–77.

Ehret, C. 2001. A historical-comparative reconstruction of Nilo-Saharan. (Sprache und Geschichte in Afrika. Beiheft 12.) Köln: Rüdiger Köppe.

Greenberg, J.H. 1963. The languages of Africa. International Journal of American Linguistics 29, 1 (Part 2). [Repr. 1966 by Bloomington: Indiana University; The Hague: Mouton.]

Gregersen, E.A. 1972. Kongo-Saharan. Journal of African Languages 11. 69–89.

Nicolai, R. 1990. Parentés linguistiques (à propos du songhay). Paris: Éditions du CNRS.

Tucker, A.N., and M.A. Bryan. 1956. The non-Bantu languages of north-eastern Africa. London, New York, Cape Town: Oxford University Press for International African Institute.

Tucker A.N., and M.A. Bryan. 1966. Linguistic analyses: the non-Bantu languages of north eastern Africa. London, New York, Cape Town: Oxford University Press for International African Institute.

MECHTHILD REH

Nootka and Wakashan Languages

Wakashan is one of a dozen or so language families indigenous to North America's Pacific Northwest, a lush and mountainous region of tremendous linguistic diversity. The term is a variation of 'Wakashian', which Captain James Cook proposed to call the people he visited in Nootka Sound in 1778:

> The word *wakash* ... was frequently in their mouths. It seemed to express applause, approbation, and friendship. For when they appeared to be satisfied, or well pleased with anything they saw, or any incident that happened, they would, with one voice, call out *wakash! wakash!*

The study of Wakashan languages has contributed significantly to the development of linguistics. For instance, the lasting and profound influence of such famous linguists as Franz Boas, Edward Sapir, Morris Swadesh, and Mary Haas is due in large part to their seminal work on these languages. Besides refuting the popular misconception that native languages are somehow primitive, their pioneering work spawned valuable research in linguistic affiliation relating to the ethnohistory of North America, and it also uncovered important structural and semantic phenomena that had not been found in more widely studied languages. Well-known contemporary linguists such as Emmon Bach and Stephen Anderson also credit their fieldwork on Wakashan languages for improving their understanding of language.

Wakashan languages fall neatly into two branches, each comprising three languages. The southern branch ('Nootkan') includes Nootka (Nuu-chah-nulth) and Nitinat (Ditidaht) spoken along the west coast of Vancouver Island, as well as Makah spoken on Washington's Olympic Peninsula. The northern branch ('Kwakiutlan') consists of Haisla-Henaksiala and Heiltsuk-Oowekyala spoken on the north and central coasts of British Columbia, and Kwakw'ala spoken on northern Vancouver Island and the adjacent mainland coast. The genetic relation between the two branches was discovered by Franz Boas in 1889. Their kinship is most clearly evidenced by the locational lexical suffixes (*-ił* 'indoors', *-as* 'outdoors on the ground', *-is* 'on the beach', *-a* 'on a rock', etc.), which are of great frequency in all Wakashan languages.

Edward Sapir and Morris Swadesh hypothesized that Wakashan belongs with Salish, Chimakuan, and Kutenai under a larger 'stock' called Mosan, and with Algic under a still larger 'phylum' called Algonkin-Wakashan (or Almosan). At present, these primeval affiliations are not widely accepted by specialists, most of whom have chosen to recoil from such large-scale classification until the histories of the smaller groupings are better understood. The lexical differences between Nootkan and Kwakiutlan alone are such that Swadesh estimated the two to have separated around three millenia ago, a time depth that is considered plausible by some (e.g. Jacobsen 1979) but too brief by others (e.g. Embleton 1985). For both branches, dialectal differentiation is greatest on Vancouver Island, which is therefore assumed to be the original home of the ancient Wakashans. The specialized proto-Wakashan vocabulary on local maritime culture also suggests that the Wakashan Urheimat lies in its present area.

Historically, the Wakashans were vigorous peoples. For instance, they used giant cedar trees to construct expansive houses for their extended families, to carve beautiful totem poles, and to build huge canoes which they used to fish for halibut and to hunt sea mammals —including colossal humpback and gray whales. But they were almost decimated by the arrival of the Europeans. Epidemics (especially smallpox), alcohol, and firearms used in intertribal wars reduced their population from tens of thousands before contact (the 1700s) to a few thousand in 1929. Fortunately, Wakashan peoples are now recovering rapidly from these historical disasters; for instance, at present, the Nootka (Nuu-chah-nulth) and the descendants of Kwakw'ala-speaking tribes (Kwakw<u>a</u>ka'wakw) number almost 8,000 and 6,000, respectively.

Unfortunately, Wakashan languages remain in a grave state of decline. The Canadian government not only prohibited Wakashan peoples' main social events ('potlatches') from 1884 to 1951 but also relocated their children to church-run residential schools from 1888 to 1983. These measures, which were taken deliberately to break down the transmission of culture and language, were all too effective. More passively, too, Wakashan languages have become obsolete under the influence of English, which has become the primary language at school and at home in every community thanks in part to the influence of the mass media. Thus, currently there are little more than a couple hundred speakers each of Nootka (Nuu-chah-nulth), Kwakw'ala, Haisla-Henaksiala, and Heiltsuk-Oowekyala. Makah has just 20 to 30 speakers, while Nitinat (Ditidaht) has less than ten. Recently, however, communities have sought to counteract the loss of their ancestral languages by integrating them in schools. To this end, they have adopted standardized orthographies, compiled dictionaries, and developed multimedia materials on their languages. The long-term effect of these efforts on the survival of Wakashan languages is unknown but hopeful.

Let us now focus on *Nootka*. This term is actually an error dating to Cook's visit. On one popular account, the legendary captain mistook it as the name of the natives who told him *nootka* ('circle about!') to come to their village on the other side of an island. The term continues to be used widely in linguistics, but nowadays the people themselves prefer to be called *Nuu-chah-nulth* ('along the mountains'). There are a dozen distinct Nuu-chah-nulth tribes, with many different dialects. Grammars exist for three of these: the southern Tseshaht, the central Ahousaht, and the northern <u>K</u>a:'yu:k't'h' (formerly Kyuquot). The following is a

brief survey of some structural characteristics of Nootka that the reader might find strikingly different and interesting.

Nootka shares 12 consonant phonemes with English, and has at least 27 others. Fifteen of these involve various kinds of glottalization. For instance, the above-mentioned dialect name *Tseshaht* is pronounced with a postglottalized 'ejective' consonant and a glottal stop [t͡s'iʃaːʔatħ], while Nuu-chah-nulth is pronounced with a preglottalized 'creaky' consonant [nuːt͡ʃaːʔnuɬ]. Nootka consonants draw on all known places of articulation except dental. For example, the dialect names <u>K</u>a:'yu:k't'h' and *Ahousaht* involve uvular, pharyngeal, and epiglottal consonants: [qaːʔjuːk'atħ, ʔaːħuːsʔatħ]. Nootka's consonantal inventory would have been even larger had it not historically abandoned the lateral approximants [l ʔl] and the voiced obstruents [b d d͡z d͡ʒ g gʷ ɢ ɢʷ], which are still in use in other Wakashan languages. In contrast to its manifold use of consonants, Nootka has just three basic vowels, in long [iː aː uː] and short [i a u] varieties, plus two marginal vowels [ɛː, ɔː] found mainly in foreign borrowings. Nootka lacks schwa [ə], which is widely used in Kwakiutlan.

Nootka phonemes are implicated in many active phonological processes, including lenition, glottalization, assimilation, shortening, and deletion. For instance, all these processes are involved in the following utterance (Tseshaht): /hiɬ-ʔaħs-uk-ʔap-maːħ/ → [hiʔjaħsukʷ'amaħ] 'I put mine in a container' (LOC-'in vessel'-POSS-CAUS-LS). Every Nootka syllable consists of a vowel obligatorily preceded by a single consonant, and optionally followed by up to three consonants, as in [.ʔintħ.tin.ʔi.] 'the one made of snot'. However, in the <u>K</u>a:'yu:k't'h dialect, vowels are regularly dropped inside and at the ends of words, such that any number of consonants is possible in sequence; for example, [t͡ɬ'uːt͡ɬ'uʃinkʷʃtɬħt͡ɬs'] (…-inuk 'at hands' + -ʃitɬ mom. + -ħts'iː 'at fire') 'he was drying his hands at the fire'. Incidentally, an extreme form of consonant sequencing occurs in Oowekyala: this Kwakiutlan dialect is remarkable in permitting consonant-only words such as [t'xt'kʷ's] 'fish hawk' and even consonant-only utterances such as [t͡ɬχspstɬkts] 'this (not visible) will be a nice thwart'. Words in Nootka show regular stress patterns (e.g. [haˈjaˌʔakˌʃiʔatɬ] 'did not know now'), as in other Wakashan languages (but not all; Haisla has a pitch accent system and Heiltsuk is tonal).

Nootka morphology is also very complex. On the one hand, it is polysynthetic in that hundreds of lexical suffixes can combine with well over a thousand roots to form large stems with composite meanings. For example, a verbal suffix like -*naːk* 'to have' can be added to a nominal root as in *t͡ʃ'apat͡s-nak* (canoe-have) 'to have a canoe', to an adjectival root as in *ʔiːħ-*

nak t͡ʃ'apat͡s (big-have canoe) 'to have a big canoe', or to a 'dummy' root as in *ʔu-naːk ʔiːħ t͡ʃ'apat͡s* (it-have big canoe) 'to have a big canoe'. It can also combine with other lexical suffixes as in *tupk-naːk-maʔiqtɬ* (black-have-want) 'to want to have a black one'. On the other hand, much of Nootka morphology is nonconcatenative in that many lexical and aspectual distinctions are expressed by reduplicating the root or by lenghtening its vowel(s). For instance, from the verb form *mitx-ʃitɬ* 'to make a turn', we get the ITERATIVE *mitxmitx-ʃitɬ* 'to start in on turns at intervals', the GRADUATIVE *miːtx-ʃitɬ* 'making a turn', the REPETITIVE *miːtxmiːtx-ʃitɬ* 'to start turning around and around', and (tenuously) the DISTRIBUTIVE REPETITIVE *mimiːtxmitx-ʃitɬ* 'to start turning around and around here and there'. In sum, the Nootka word consists of a root, which itself may be modified by reduplication or lengthening, and to which may be added a large number of suffixes with various lexical and grammatical meanings.

Turning to syntax, it is often obscured in Nootka by the complex morphology that allows words to convey sentences, as in <u>K</u>a:'yu:k't'h' *ʔu-kʷiːɬ-t͡ʃip-'it͡ʃ-is-im* (it-make-for-IMPV-you(PL)-me-will) 'you'll make it for me!', *ʔu-suːp-intiʃ* (it-kill-PST.IND) 'he killed it', and *muk-suːp-intiʃ* (deer-kill-PST.IND) 'he killed the deer'. When individual words are used instead of affixes, the following basic order is apparent: predicate–subject–object, as in *qaħsaːp-intiʃ t͡ʃakup muwit͡ʃ* (kill-PST.IND man deer) 'a man killed a deer'. Interestingly, any category of word can be a predicate in Nootka, including adverbs as in *ʔaħti-intiʃ qaħsaːp muwit͡ʃ* (nightly-PST.IND kill deer) 'it was at night he killed the deer', nouns as in *t͡ʃakup-intiʃ qaħsaːp muwit͡ʃ* (man-PST.IND kill deer) 'it was a man that killed the deer', and demonstratives as in *ʔuħ-intiʃ t͡ʃakupi qaħsaːp muwit͡ʃ* (that-PST.IND man kill deer) 'it was that man that killed the deer'. Of particular interest in this context is the oft-repeated claim that Nootka lacks category distinctions. This claim is controversial, but most linguists agree that this kind of distinction is weak in Nootkan syntax. For example, in Tseshaht *quːʔas* 'man' and *mamuːk* 'work' act as noun and verb, respectively, in *mamuːk-ma quːʔas-ʔi* 'the man is working' (work-he man-the), but these grammatical categories appear to be reversed in *quːʔas-ma mamuːk-ʔi* 'the working one is a man' (man-he work-the).

References

Boas, Franz. 1947. Kwakiutl grammar. Philadelphia: American Philosophical Society.

Cook, Captain James. 1784. A voyage to the Pacific Ocean: undertaken by the Command of His Majesty, for making discoveries in the Northern Hemisphere; performed under the direction of Captain Cook, Clerke, and Gore, in His

Majesty's Ships the Resolution and Discovery, in the years 1776, 1777, 1778 and 1780. London: W. and A. Strahan.

Embleton, Sheila M. 1985. Lexicostatistics applied to the Germanic, Romance, and Wakashan families Word 36. 37–60.

Goddard, Ives (ed.) 1996. Languages. Handbook of North American Indians, Vol. 17. Washington: Smithsonian Institution.

Haas, Mary R. 1972. The structure of stems and roots in Nootka–Nitinat. International Journal of American Linguistics 38. 83–92.

Jacobsen, William H. Jr. 1979. Wakashan comparative studies. The languages of North America: historical and comparative assessment, ed. by Lyle Campbell and Marianne Mithun. Austin, TX: University of Texas Press.

Lincoln, Neville J., and John C. Rath. 1980. North Wakashan comparative root list. Ottawa: National Museums of Canada.

McMillan, Alan D. 1999. Since the time of the transformers: the ancient heritage of the Nuu-chah-nulth, Ditidaht, and Makah. Vancouver: University of British Columbia Press.

Mithun, Marianne. 2001. The languages of Native North America. Cambridge, UK: Cambridge University Press.

Nakayama, Toshihide. 1999. Nuu-chah-nulth (Nootka). Munich: LINCOM EUROPA.

Sapir, Edward, and Morris Swadesh. 1939. Nootka texts: tales and ethnological narratives with grammatical notes and Lexical materials. Philadelphia: American Philosophical Society.

Stonham, John T. 1999. Aspects of Tsishaath Nootka phonetics and phonology. Munich: LINCOM EUROPA.

DARIN HOWE

Northwest Caucasian Languages

The languages of the Northwest Caucasian family—Abaza, Abkhaz, Adyghe, Kabardian, and Ubykh—were originally spoken in a contiguous region stretching from Abkhazia on the Black Sea in the south to the Kuban River in the north. As a result of the relocations following the Russo-Caucasian War (1817–1864), substantial communities of each now exist in Turkey, as well as smaller enclaves in Syria, Jordan, Israel, former Yugoslavia, Western Europe, and New Jersey. The languages of this family fall into three groups: Abkhaz/Abaza, Ubykh, and Circassian, the latter containing the languages Adyghe and Kabardian.

In 1850, there were approximately one million speakers of Northwest Caucasian languages, most of them Circassians. The Russo-Caucasian war and the Russian colonization of the Caucasus drastically changed this situation: many Abkhaz and Circassians were forced to flee westward to the Ottoman Empire, and the Ubykh and Sadz Abkhaz were entirely displaced from their homelands near the Kwydypsta River just north of present-day Abkhazia. At present, some 800,000 Northwest Caucasians still reside in the Caucasus, most of whom retain their ancestral languages, as well as Russian. In addition, more than a million Northwest Caucasians now live in the diaspora, primarily in Turkey. Current estimates of the number of North Caucasians in Turkey range from one to six million, depending on the sources used. It is difficult to generate reliable figures, because many Northwest Caucasians now have complicated ethnic identities resulting from cultural and linguistic assimilation and intermarriage. In the Northwest Caucasus

proper, the different ethnic groups have generally preserved their own languages, although for some Russian has become the dominant language.

Many Caucasian linguists believe that the Northwest Caucasian family is genetically related to the neighboring Northeast Caucasian family, but the linguistic arguments presented in support of this connection are not convincing, resting on typological similarity rather than the comparative method. All of the Northwest Caucasian languages have literary forms, with the exception of Ubykh, which is now extinct. Their current orthographic forms are based on the Cyrillic script, although there is currently movement to develop a new orthography based on the Roman script (most likely its Turkish manifestation). Abkhaz and Abaza are quite similar to one another and are considered by some scholars to be dialects of the same language; the same holds for Adyghe with respect to Kabardian. Ubykh is linguistically intermediate between these two groups.

The Northwest Caucasian languages are most famous for pairing unusually rich consonant inventories (containing as many as 83 consonants, in Ubykh) with unusually small vowel inventories, typically consisting (in native words) of only a low vowel /a/ and a non-low vowel /ɨ/. The richness of the consonant system derives for the most part from the addition of labial and palatal secondary articulations. It appears that the enormity and smallness of the consonantal and vocalic systems, respectively, are related: the ancestral (proto) language most likely had more conventional vowel and consonant systems, and subsequently reinterpreted the [back] and [round] features of the vowels

as resulting from secondary7 articulations on neighboring consonants. (Turkish, Arabic, and Persian but not Russian loans underwent this reanalysis as well: cf. Turkish dükkân 'store'→Abkhaz /a-dʷɨk'yan/ [ádˌbukʼʲán].) One can see in this example that what are now phonologically consonantal secondary articulations still surface on the vowels in most contexts. All of the Northwest Caucasian languages distinguish three types of laryngeal consonants, opposing plain voiced, voiceless aspirated, and ejective stops. Many also possess extremely complicated intonational stress systems, most famously Abkhaz.

The system of agreement marking in Northwest Caucasian verbs is also uncommonly intricate. Abkhaz, for example, overtly marks verbal agreement with subjects, direct objects, indirect objects, and various other oblique arguments, as in the following Abzhuy example from George Hewitt (transcribed in the IPA):

saɾá	ɑ-pħʸɨs	ɑ-sɑpʼɨn
I	the-woman	the-soap

s-χaɾpʰ (∅-)ɑ-lɑ-l sɨ-ɾ-dʒʸɨdʒʸá-(∅-)jtʼ
my-shirt (it-)it-by her-I-cause-wash-(PAST-)FINITE
'I got the woman to wash my shirt with (the) soap'

It is possible to string together as many as nine overt morphemes in a row in a sequence of verbal prefixes.

The nominal morphology, on the other hand, is relatively simple, containing only two cases (direct object and oblique or adverbial). Possession is expressed via prefixes, as in Abkhaz sɨ-napʼ-kʷá 'my hands' (I-hand-nonhuman.plural). Abkhaz (and Abaza) distinguish human vs. nonhuman in certain morphological classes, as can be seen in the plural in this example. All of the Northwest Caucasian languages distinguish two genders, two numbers, and three persons in their system of agreement markers.

Finding lexical items shared by all of the Northwest Caucasian languages is difficult, but not impossible:

Proto-Northwest Caucasian	Common Circassian	Ubykh	Common Abkhaz
*pxʲa 'back'	pxa	pšʲa	*(p)χʲa
*bza 'tongue'	*bza	bzʲa	*bzə
*zʷʲa 'cook, boil'	*zʷʲa	zʷʲa	*zʷʲə
*gʷʲə 'heart'	*gʷə	gʲə	*gʷə

The Northwest Caucasian numeral system is interesting in that it preserves vestiges of an original vigesimal system in the numbers 30–99. For instance, Abkhaz (Cwyzhy dialect) eцχʷejцχʷɑbɑ 'thirty' is literally 'twenty (eцχʷɑ) and (-j) ten (zʷɑ-)'; цɨneцχʷɑ 'forty' is 'two twenties', and so on.

Turning to the individual languages, Abaza had 34,800 speakers in the Karachay-Cherkess autonomous region of Russia at the time of the 1989 census; as of 1995, there were 10,000 more in Turkey, 80 (out of 150 members of the ethnic group) in Germany, and about 15 in the United States. There are three main dialects: Tapanta, Ashkhar, and Bezshagh. The literary language is based on the Tapanta dialect. Abaza is reported to be mutual with Abkhaz.

Abkhaz had 101,000 speakers in Abkhazia in 1993, as well as 4,000 speakers out of approximately 15–30,000 ethnic Abkhaz in Turkey, and smaller numbers scattered in other countries. Abkhazia was under Georgian rule during the Soviet period, but seceded shortly after the dissolution of the Soviet Union in the mid-1990s. The continued refusal by foreign nations to recognize Abkhazia's sovereign status, together with the forced immigration into Abkhazia of Slavs, Armenians, and Mingrelians during the 1930s, and Georgia's continuing attempts to eliminate its ethnic minorities, have placed the Abkhaz language in danger of disappearing. The assignment of literary status to Abkhaz (including TV broadcasts since 1978) has helped counteract this tendency within Abkhazia, but few members of the younger generation are learning the language in the diaspora. There are three main dialect groups: Bzyp (north of Sukhumi), Abzhui, the literary dialect (south of Sukhumi), and Sadz (now spoken only by a handful of Abkhaz in Turkey).

The Circassians live in the Adyghe, Kabardino-Balkar, and Karachay-Cherkess republics in the Russian Federation. The Circassians have two main languages, Adyghe (West Circassian) and Kabardian (East Circassian), the former of which has four main dialects (Shapsugh, Bzhedug, Temirgoy (the literary dialect), and Abzakh) and the latter of which has six (Besleney, Malka, Bakhsan, Terek, Mozdok, and Lesser Kabardian). As of 1993, there were some 125,000 Adyghe and 46,000 Kabardians in Russia, 100,000 Adyghe and 202,000 Kabardians in Turkey, and a total of approximately 750,000 Circassians in the world as a whole. Both languages have literary status and are taught in the schools in their respective republics, but their survival in the diaspora is less secure following the closing of the Circassian school in Turkey by Atatürk in 1920 and the outlawing of publication in minority languages in Turkey in 1983.

The Ubykh community originally lived on the Black Sea just north of Abkhazia, but all 50,000 of its members were deported to the Ottoman Empire following the Russian conquest of the Caucasus, ending up in Hacı Osman, a town near Istanbul. Although the community retains a distinct identity, its language is now extinct. Isolated individuals can still be found who know pieces of the grammar and vocabulary, but

most of the remaining Ubykhs speak Adyghe and/or Turkish.

References

Allen, Sidney. 1956. Structure and system in the Abaza verbal complex. Transactions of the Philological Society. 127–76.

Chirikba, Viacheslav. 1996. Common West Caucasian. Leiden: CNWS Publications.

Colarusso, John. 1988. The Northwest Caucasian languages: a phonological survey. New York: Garland Press.

Colarusso, John. 1992. A grammar of the Kabardian language. Calgary: University of Calgary Press.

Dumézil, Georges. 1931. La langue des oubykhs. Paris: Champion.

Dumézil, Georges. 1932. Études comparatives sur les langues caucasiennes du nord-ouest (Morphologie). Paris: Adrien-Maisonneuve.

Hewitt, George. 1979. Abkhaz. Amsterdam: North-Holland.

Hewitt, George (ed.) 1989. The indigenous languages of the Caucasus, Vol. 2: The North West Caucasian languages. Delmar, NY: Caravan Books.

Kuipers, Aert. 1960. Phoneme and morpheme in Kabardian. The Hague: Mouton.

Nikolaev, S., and S. Starostin. 1994. A North Caucasian etymological dictionary. Moscow: Asterisk Publishers.

Paris, Catherine. 1974. Système phonologique et phénomènes phonétiques dans le parler besney de Zennen Köyü. Paris: Klincksieck.

Smeets, Rieks. 1984. Studies in West Circassian phonology and morphology. Leiden: Hakuchi Press.

Sommerfelt, Alf. 1934. Études comparatives sur le caucasique du Nord-ouest. Norsk Tidsskrift for Sprogvidenskap 7. 178–210.

von Mészáros, Julius. 1934. Die Päkhy-Sprache. Chicago: University of Chicago Press.

BERT VAUX

Noun Incorporation

The term Noun Incorporation refers to the linguistic phenomenon according to which a noun and a verb may combine into a single word. The two elements form one morphological unit that is often described as a complex verb (or as a compound). Usually, the incorporated noun is interpreted as the object of the verb.

For example, in Onondaga, an American Indian language of the Iroquian language family, the noun (stem) *hwist* 'money', which is in **bold** below, may appear as part of the verb (stem) *ahtu* 'lose'. In particular, the noun (stem) immediately precedes the verb (stem), as we can see in the following example:

(a). Pet   wa?-ha-**hwist**-ahtu-?t-a?
     Pat   he-**money**-lost
     'Pat lost money'

In this case, the noun (stem) and the verb (stem) form a single word, namely the complex verb *wa?-ha-hwist-ahtu-?t-a?* 'he-**money**-lost'. The incorporated noun (stem) *hwist* 'money' is interpreted as the object of the verb *ahtu* 'lose'.

In the same language, the object of the verb can be an independent word in the sentence. In this case, the object is one word: *ne?o-hwist-a?* 'the money', while the verb is another word, namely, *wa?-ha-htu-?t-a?* 'he-lost'. The noun now follows the verb, as illustrated below:

(b)   Pet   wa?-ha-htu-?t-a?   ne?o-**hwist**-a?
      Pat   he-lost   the-money
      'Pat lost the money'

(Examples (a) and (b) are taken from Baker 1988:76.)

Incorporating languages thus have two ways of expressing the verb and its object: either the two elements are realized as a single word after incorporation of the noun into the verb, as shown in example (a), or they appear as two separate words. In the latter case, the noun does not incorporate into the verb, but it follows the verb, as shown in example (b). The two constructions may differ slightly in their interpretation. For example, Incorporation sometimes has the function of narrowing the semantic range of the verb. This implies that the speakers of incorporating languages chose between the nonanalytic configuration (where the object is incorporated into the verb) and the analytic configuration (where the object is not incorporated into the verb) on the basis of the particulars of the information they want to communicate.

Noun Incorporation is a productive process mainly in non-Indo-European languages. It is attested in some North American languages (Athapaskan, Caddoan, Chimariko, Iroquian, Siouan, Kiowa-Tanoan, Natchez, Takelma, Uto-Aztecan, Yana, Zuni), in Australian languages (Rembarnga), in Paleo-Siberian languages (Chukchee), in Oceanic languages, in Turkish, and in other languages.

There has been a long debate in the literature of Generative Grammar regarding Noun Incorporation. The debate concerns the locus where these complex words are formed. On the one hand, it has been argued

that Noun Incorporation is an instance of compounding in the Lexicon (Mithun 1984; Di Sciullo and Williams 1987; Rosen 1989). Compounding refers to the combination of two (or more) stems or words into one single word. For example, the English word *bedroom* is a compound consisting of two elements: *bed* and *room*. Hence, it has been argued that Noun Incorporation is a case of a noun and a verb that are put together in the same way that *bed* and *room* are combined into *bedroom*. According to this view, the noun–verb complex is not related structurally to its analytic counterpart. That is, there is no association between the noun–verb compound and the structure that involves the verb and its object, despite the fact that the two configurations may have very similar interpretation. The noun and the verb are put together morphologically and behave as one syntactic unit. This hypothesis is usually known as the lexicalist hypothesis.

Baker (1988), on the other hand, was the first to argue that Noun Incorporation is a syntactic phenomenon. His main idea is that Noun Incorporation is a syntactic process that derives complex verbs from sentences where the verb takes a noun as its object. Specifically, Baker argues that the incorporated noun starts out as an independent word, in a position following the verb. It ends up in a position preceding the verb after an operation of syntactic movement. That is, syntactic movement changes the position of the noun and places it within the verbal complex. An empty slot is left behind, in the original postverbal position of the noun, which is called the trace of the moved element.

Baker provides a straightforward explanation of the fact that the incorporated noun is interpreted as the object of the verb. It starts out as the complement of the verb, i.e. the two elements are adjacent and in a specific order: the noun follows the verb. This is considered as the structural position where the internal theta-role of the verb is assigned (following Baker's Universal Theta-role Assignment Hypothesis—the UTAH—according to which each thematic role of the verb is assigned in a specific syntactic configuration). That is, any element that is base-generated in that particular syntactic position is interpreted as the direct object of the verb, namely as the theme (i.e. the argument that is affected by the action described by the verb). Hence, if the incorporated noun starts out as the complement of the verb, it is assigned the internal theta-role of the verb and is thus interpreted as the object of the verb, even after incorporating (by syntactic movement) into the verb. Incorporation is predicted to be possible with any noun that starts out from the verbal complement position. This is actually true: Noun Incorporation is a productive process in incorporating languages.

In nonincorporating languages, like English, the object cannot form a single word with the verb. Words like *truck drive* are not in use. This means that the noun

truck, which is the object of the verb, cannot become part of the verb *drive*, but rather the two elements have to remain two separate words. The noun *truck* can be combined with the noun *driver,* giving rise to a complex noun: *truck driver*. However, it can be argued that these cases are different from genuine instances of Noun Incorporation in that the derived word is a noun rather than a verb. There are some complex words in English that resemble cases of Noun Incorporation. It could be argued that the formation of verbs like *baby sit* possibly involves an operation of syntactic movement. However, such examples lack a parallel analytic configuration. That is, *to sit a baby* cannot be considered as the parallel analytic configuration of the complex verb to *baby sit*. The latter has a very unexpected meaning. It expresses the notion 'to look after a baby'. If *baby sit* is not an instance of Noun Incorporation, it could be viewed as an instance of compounding. This would mean that the verb *baby sit* is derived by a lexical or a morphological process and not by a syntactic process.

Lastly, it has been argued that Incorporation is a widely spread phenomenon that is not necessarily restricted to nouns. It is a process that allows any two elements to combine into a single word. One of the two parts of the derived word is usually a verb. It has been proposed, for example, that certain kinds of pronouns incorporate into the verb. These are called clitic pronouns and are set apart from other pronouns by certain special characteristics. While ordinary pronouns do not appear as parts of the verb, clitic pronouns may do so. This process is often named cliticization. The process of cliticization can be distinguished, though, from Noun Incorporation in that the former possibly involves incorporation of a functional element (a clitic pronoun) whereas the latter involves incorporation of a lexical element (a noun).

References

Baker, Mark. 1988. Incorporation. Chicago: University of Chicago Press.

Baker, Mark. 1996. The polysynthesis parameter. Oxford: Oxford University Press.

Di Sciullo, Anna Maria, and Edwin Williams. 1987. On the definition of word. Cambridge, MA: MIT Press.

Mithun, Marianne. 1999. The languages of native North America. Cambridge: Cambridge University Press.

Mithun, Marianne. 1984. The evolution of Noun Incorporation. Language 60. 847–94.

Rosen, Sara Thomas. 1989. Two types of Noun Incorporation: a lexical analysis. Language 65. 294–317.

Spencer, Andrew. 1991. Morphological theory. Oxford: Blackwell.

Spencer, Andrew, and Arnold Zwicky (eds.) 1998. The handbook of morphology. Oxford: Blackwell Publishers.

DIMITRA PAPANGELI

See also **Generative Grammar; Mohawk and Iroquoian Languages**

Number Marking

Number is the morphological category that expresses contrasts involving countable quantities. The most widespread number opposition is likely that between singular (one) and plural (more than one). Several languages, however, make further distinctions. Some of them (e.g. Greek or Sanskrit) exhibit dual forms that apply to two individuals, and a few languages (e.g. some Southwest Pacific languages) additionally distinguish a trial form that refers to three individuals. In some systems, there are also paucal forms for indicating a small number, as in Arabic.

Although number implies a basic linguistic contrast, languages deal with this category in different ways. Most languages mark number in nouns; yet, there are languages (e.g. Maori, a Polynesian language) in which it is signaled by determiners, and others that lack number distinctions (e.g. Nancowry, spoken in India's Nicobarese Islands). The explicit expression of number, on the other hand, may be obligatory, as in most English nouns; limited to certain nominal categories, as in most Algonquian languages (spoken by American Indians); or optional, as in Halkomelen (another American Indian language).

Number marking differs greatly among languages, depending on the morphological system they display. In languages traditionally classified as analytic, such as Chinese, words consist of only one syllable and are invariable. Thus, plural nouns do not exhibit any formal marker and the notion of plurality must be inferred from the context. Only when this information is not contextually available or the idea of plurality is to be emphasized, separate words are added: *ki* 'some', *šu* 'number'. Synthetic languages, on the contrary, rely on inflection, i.e. the modification of a word's form, to express number distinctions. Modifications of words may equally result in derivation, which involves the creation of new words with new meanings. In traditional morphological theory, which has its origins in American structuralism, the analysis of word structure was based on the morpheme—the smallest meaningful unit of language. Morphemes are, however, abstract elements that are physically realized by morphs. The word *houses*, for example, consists of two morphs: *house* and *-s*. It is very frequent that a particular morpheme is represented by different morphs in different environments. For instance, the plural morpheme in the English noun *oxen* is not realized by the regular morph *-s*, but by the irregular *-en*. These alternants are called allomorphs.

Both inflection and derivation are traditionally explained exclusively in terms of affixation. As regards inflection, affixation consists in the addition of an inflectional affix (an obligatorily bound morph that contains grammatical information) to another morph with lexical content, as in *house-s*. This approach to inflection, however, is said to face important difficulties. One of them is illustrated by irregular English plurals such as *men,* in which plurality is marked not by the addition of an allomorph, but rather by means of a process that causes a vocalic change. The so-called portmanteau morphs (morphs that realize more than one morpheme, as the ending *-us* in the Latin noun *annus* 'year', which indicates simultaneously singular number and nominative case) are also problematic, since this approach assumes a one-to-one correspondence between form and meaning. In view of such shortcomings, recent morphological theories of inflection have abandoned this classical treatment and have adopted a model in which words are modified by means of different morphological processes that apply to the word itself, which is considered the base or stem.

The most simple device to express number is, as in traditional accounts, affixation. Affixes can be divided into three main classes: prefixes, which attach to the front of the stem; suffixes, which come after the stem; and infixes, which occur within the stem. Most Indo-European languages make use of suffixes, such as the English or Spanish regular plural ending *-s*, to express number contrasts. Number marking by means of prefixation can be found in Bantu languages such as Kikuyu (*mũ-rũ ũ thi* 'lion', pl. *mĩ-rũ ũ thi*), whereas the Uto-Aztecan language Oaxaca Chontal illustrates the use of infixes to distinguish number (*kwepoʔ* 'lizard', pl. *kwe-ł-poʔ*).

Other processes by which number oppositions may be expressed involve the modification of the stem. This shift may affect the quantity or quality of the internal vowel, as in the English pair *foot/feet*. Vowel change is generally referred to as 'ablaut', although, because of its historical origin, the alternation just mentioned receives the distinct name 'umlaut'. Not only vowels but also consonants may be modified to mark number. In the West African language, Fula, the plural of the word *yiite* 'fire', is formed by changing the initial consonant, *giite*. Such consonant alternations are frequently the result of a phonological change induced by an affix, which is then lost.

In many languages, information about number is conveyed by means of modifications of the suprasegmental features tone and stress. In Somali (an East African language), for example, number is indicated by a change in the tone pattern: the singular form *èy* 'dog' takes a falling tone, whereas the plural *éy* takes a high tone. Russian illustrates the use of stress shift to distinguish singular from plural: in a particular group of neuter nouns, the singular is ending-stressed, while the plural is stem-stressed: *oknó* 'window' (nominative singular), *ókna* (nominative plural).

Another common morphological process in certain languages is reduplication, which consists in the copying of the whole stem—total reduplication—or only of a part of it—partial reduplication, typically the leftmost portion. Reduplication is said to be frequently used iconally, i.e. the form of the word reflects its meaning. For this reason, reduplication is a common marker of plurality. Total reduplication to form the plural is found in Indonesian (*babi* 'pig', pl. *babibabi*), whereas partial reduplication is used in Motu, a language of Papua New Guinea (*tau* 'man', pl. *tatau*).

The last morphological operation associated with number marking is suppletion. This process replaces one form by a phonologically unrelated form. A clear example of this phenomenon is the use of *went* as the past tense of *go*. Suppletion appears in many Western North American languages as a number marking device: in Navajo, the form for 'one is standing' is different from that for 'several are standing'.

References

Anderson, Stephen R. 1985. Inflectional morphology. Language typology and syntactic description, Vol. III, ed. by Timothy Shopen. Cambridge: Cambridge University Press.

Anderson, Stephen R. 1988. Morphological theory. Linguistics: the Cambridge survey, Vol. I, ed. by Frederick J. Newmeyer. Cambridge: Cambridge University Press.

Anderson, Stephen R. 1992. A-morphous morphology. Cambridge: Cambridge University Press.

Bauer, Laurie. 1988. Introducing linguistic morphology. Edinburgh: Edinburgh University Press.

Bloomfield, Leonard. 1933. Language. New York: Holt, Rinehart, and Winston; Chicago and London: The University of Chicago Press, 1984.

Jespersen, Otto. 1922. Language. Its nature, development and origin. London: George Allen & Unwind Ltd.

Matthews, P. H. 1972. Inflectional morphology. A theoretical study based on aspects of Latin verb conjugation. Cambridge: Cambridge University Press.

Matthews, P. H. 1974. Morphology. An introduction to the theory of word-structure. Cambridge: Cambridge University Press.

O'Grady, William, Michael Dobrovolsky, and Francis Katamba (eds.) 1987. Contemporary linguistics. An introduction. London and New York: Longman; 3rd edition, 1996.

Spencer, Andrew, and Arnold M. Zwicky. 1998. The handbook of morphology. Oxford, Malden, MA: Blackwell.

LUISA GONZÁLEZ-ROMERO

See also **Affixation; Inflection and Derivation**
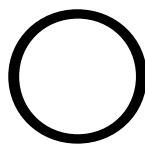
Official Language Selection

Official language selection may be simply defined as the political choice of which language, or languages, to use in the legislative, executive, and judicial business of government. Given that all nations in the twenty-first century comprise populations speaking different languages, the selection of official languages is much more complex than the definition. This is because granting official status to a particular language will enhance its prestige, extend its use to educational and nonofficial domains, privilege its speakers, and impinge on the linguistic rights of speakers of other languages within the community. Both newly independent and well-established countries face difficulties when selecting official languages, although the contributing factors differ. Examples of each will now be considered.

Countries that are colonized or subjected to imperialism in one form or another usually have the language of the colonial state imposed on them. Upon reaching independence, the selection of the official language is a matter of considerable importance—and controversy—because there are often several contending languages. One of them may be the first language of a majority of, or dominant group within, the population, and this would seem the obvious choice for nation building. This was the case with the nations that re emerged after the demise of the Soviet Union in the 1990s (Ukraine, Latvia, Belarus, and so on). However, in such cases, speakers of other indigenous languages may feel that they are regarded as second-class citizens for whom access to public services is restricted, if not denied completely, by the lack of official recognition of their mother tongue. This situation arose in Pakistan, where Urdu was declared the official language on independence from Britain in 1947. However, because Urdu was not an indigenous language of East Pakistan, the overwhelmingly Bengali-speaking majority there felt marginalized; the unrequited demand for language rights led to civil war, after which Bangladesh achieved independence from Pakistan in 1971. Therefore, some newly independent nations select the language of the previous imperial power; this was the pattern among most ex-colonies of Britain in the 1960s and 1970s, such as Ghana, Nigeria, and Kenya. Although this choice means that a new nation may relatively easily build on the existing communications infrastructure, the choice may be inappropriate. There is often a residue of ill-feeling toward the culture and language of the erstwhile rulers; such a decision may also leave in positions of influence those who collaborated with the colonial power and who may be out of sympathy with a new regime. Because of this, some countries adopt another international language. The example of East Timor, which became independent in 2000, is a case in point. None of the indigenous languages was considered appropriate as the official language. The language of the erstwhile dominant power—Indonesia—was considered unacceptable on political and cultural grounds. Portuguese—the language of the earlier colonizer—was eventually chosen because it was widely understood and was used both within East Timor and internationally. Since 1994, South Africa has presented a very different example of language policy—one of national pluralism. At the national level, 11 indigenous languages have official status, and each provincial

legislature can determine which of the languages will be used for internal official purposes.

Longer established nations may also have difficulty in selecting official languages, as examples from major English-speaking countries illustrate. For many years, the francophone population of Canada struggled for legal recognition of their language until eventually the 1969 Official Languages Act declared that English and French had equality of status with equal rights and privileges. The case of the United Kingdom is different because although English is the national language, it has no official status: the only legally recognized official language in Great Britain is, since 1967, Welsh. Similarly, since 1987, the only official language throughout New Zealand is Maori. The reason for this apparent legal anomaly is that the indigenous speakers of those two minority ethnic groups campaigned to have their languages officially sanctioned for use in public services such as the legal system and education, health, and social services. With regard to English, there was no need to formalize such rights because they existed in fact, although not in law. Some English-speaking people in these countries opposed the granting of prestige and rights to these minority languages; they also opposed the allocation of public resources for their development and diffusion. None of these countries has an explicit overall national policy on languages, and there is considerable scope for uncertainty and even anxiety regarding the language rights of speakers of nonofficial languages:for example, immigrants and minority ethnic communities in all three countries and the indigenous First Nation peoples of Canada. By contrast, Australia has been developing a comprehensive National Languages Policy since 1987; although this now declares that English is the official language of Australia, it recognizes the aboriginal languages and also specifies the provision of services in languages other than English. The United States has never selected an official language, despite the large number of indigenous, colonial, and immigrant language communities. Recent attempts to change the federal constitution to adopt English as the official language throughout the United States have been unsuccessful, although since the 1980s many state legislatures have passed laws proclaiming, with varying degrees of specificity, English as the official language within their own borders.

Language is not merely a means of communication: it is the most important cultural symbol of a community. For many countries—Japan, Somalia, Malta, and many others—the national language is a core value that bonds the people together and indeed defines that very community. Increasingly, however, all nation states are host to diverse linguistic minority communities in their midst, whether this is caused by temporary business activity, permanent migration, urgent asylum-seeking, or the revitalization of indigenous languages. It is, therefore, necessary for national policy-makers to consider the extent to which multilingualism (and multiculturalism) presents either a threat or a resource to their communities, and thereafter to select official languages and frame national language policies in the light of this understanding.

References

Bourdieu, P. 1991. Language and symbolic power. Cambridge: Harvard University Press.

Eggington, William, and Helen Wren. 1997. Language policy: dominant English, pluralist challenges. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Fairclough, N. 1989. Language and power. London: Longman.

Herriman, Michael, and Barbara Burnaby (eds.) 1996. Language policies in English-dominant countries. Clevedon, Philadelphia, Adelaide: Multilingual Matters Ltd.

Phillipson, Robert. 1992. Linguistic imperialism. Oxford: Oxford University Press.

Schiffman, H.F. 1995. Linguistic culture and language policy. London: Routledge.

Tollefson, James, W. 1991. Planning language, planning inequality. London, New York. Longman.

Varennes, Fernand de. 1996. Language, minorities and human rights. The Hague: M. Nijhoff.

ROGER BARNARD

See also **Australia; Canada; Language Planning; Nigeria; South Africa; Soviet Union: Successor States**

Ojibwe and Algonquian Languages

Ojibwe (sometimes Ojibwa, Chippewa) is a Native American language and belongs to the Algonquian family, one of the largest of precontact North America. Although widely dispersed, these languages share many of the same grammatical features. Other members in this group include Blackfoot, Cheyenne and

Arapaho, Cree, Potawotami, Menominee, Fox, Illinois, Shawnee, Micmac, Passamaquoddy, and Abenaki.

Algonquian speakers have inhabited North America for thousands of years. Their homeland was suggested to be somewhere in the region of Lake Huron or Georgian Bay, but given the ease of movement throughout the eastern woodlands area, most researchers are noncommittal in this regard. At the time of European contact, Algonquian languages were spoken along the eastern seaboard north of Virginia, as well as to the east of the lower Mississippi River. Today they can be found from the Atlantic Ocean to the Rocky Mountains along what is now the United States/Canadian border. Place names such as Connecticut, Manitoba, Massachusetts, Michigan, Minnesota, Mississippi, Ontario, Saskatchewan, and Wisconsin are all Algonquian in origin, as are the words *moose*, *chipmunk,* and *toboggan.*

Linguists often speak of Algonquian in terms of regional subdivisions, but only Eastern Algonquian constitutes a family in the true sense; i.e. it can be shown that these languages have a common ancestor. Members of this group include Abenaki (spoken in Québec and Maine), Passamaquoddy-Maliseet (Maine, New Brunswick), and Micmac (Nova Scotia, New Brunswick). Arapaho, Cheyenne (Wyoming), and Blackfoot (Montana, Alberta) fall into the western group. 'Central Algonquian'—of which Ojibwe is a member—is spoken around the area of the Great Lakes. Ottawa (Odawa), Algonquin, and Salteaux (all spoken in Canada) are essentially the same as Ojibwe; Cree, Fox, and Shawnee are close relatives. Divisions along geographical lines are not completely arbitrary: members of each region share certain sounds, vocabulary, and sentence patterns not found in others. The differences between one language and the next shift gradually from west to east, with western languages being the oldest.

Algonquian was one of the first language families to be reconstructed by using methods developed in the nineteenth century for Indo-European. By comparing words in various languages, the American structuralist Leonard Bloomfield was able to determine the forms of 'proto' Algonquian, at the same time making significant contributions to our understanding of linguistic change itself. Since then, many other researchers have followed in Bloomfield's footsteps, to the point where Algonquian is often looked on as a textbook case of structuralist analysis. Outside of taxonomy and analyses of sound change, not much systematic work has been done on Algonquian. Compared with other languages, our understanding of word and sentence structure is incomplete, and research on meaning and discourse is still in its infancy.

As a spoken language, Ojibwe is fairly robust, whereas several other Algonquian languages have declined or died out altogether. Most speakers live on reservations in Minnesota, Wisconsin, and Michigan (United States), as well as in the province of Ontario (Canada). One recent estimate puts the total number of Ojibwe speakers at 51,000, but most other estimates are lower. In any case, only a small percentage represents unilingual Ojibwe speakers. Many efforts are under way to preserve and/or revive the language. These usually take the form of bilingual education programs in the schools, language centers for adults, or the establishment of institutional or personal websites. Recently published word lists and dictionaries abound, and Ojibwe has received a relatively good deal of attention from linguists. Otherwise, the model of loss is similar to endangered languages worldwide: younger speakers are heavily influenced by the dominant culture (in this case, Anglo-America) and use their first language primarily in family or tribal situations. Borrowed speech habits continually erode native sound and sentence patterns until only the elders speak the language fluently.

The sounds of Ojibwe are typical of other Algonquian languages. There are five vowels of different quality, and vowel length (indicated by a colon) distinguishes meaning. The vowels are [i, e, (schwa), i:, a:, and o:]. The consonants are [p, b, g, k, m, n, ch, j, w, y ,h, and ʔ]. Note the absence of [t, d].

Ojibwe (henceforth Algonquian) is a polysynthetic language, a term that carries with it many implications: single, complex words typically express the meaning of whole sentences, pronouns are 'understood' rather than overt, and expressions such as 'all' or 'every' may be separated from the nouns they modify by other words. In addition, polysynthetic languages are rich in agreement: subjects, direct objects, and even indirect objects cause the verb to inflect for person, number, and gender—and sometimes case. Noun phrases, meanwhile, do not appear in fixed positions within the sentence, as in English; instead they are considered 'appositional' (add-ons), only loosely linked to the verbs. Many meanings that are conveyed in full phrases in other languages are expressed via particles that attach to the verb.

Verbs in Ojibwe (and all Algonquian languages) agree with subjects as well as direct objects, sometimes in very intricate ways. In the sentence *[ni-ki:-wa:pem-a:-n] kene:peko:n* '[I saw] the snake', the prefix *ni-* indicates that the subject is first person, whereas the absence of a plural marker ensures that it is singular. The third-person feature of *kene:peko:n* 'snake' is encoded by the suffix *-a:,* whereas *-n* signals that the (singular) object is different from the subject (cf. obviation below). Because of such rich agreement, there is no need for separate pronouns other than

emphatic ones. In the absence of an object, this sentence would simply mean 'I saw him/her/it.' When plural morphemes are added to the verb, agreement is 'skewed', yielding the order

person prefix -	person -	number -	number
(verb stem)	suffix	suffix	suffix
SUBJECT	OBJECT	SUBJECT	OBJECT

In other words, subject as well as object agreement is split between two nonadjacent positions. Various linguistic theories have attempted to deal with facts like these, but so far very few have succeeded in providing a natural account. Finally, nouns inflect for person, number, and obviation too, as in *ni-gwis-ag* [first person-son-plural] 'my sons'. Note that often the same affixes involved in verbal agreement also show up in this context.

A large class of elements called preverbs appear in fixed positions before the verb stem. These typically modify the state or action denoted by verb, much like modifier phrases do in English. Examples from Ojibwe are *ojaanimi* 'busy, noisy', *wani* 'mistakenly', and *aano* 'in vain, without result'. Some preverbs are themselves verbal in character, such as *giizhii* 'finish, get through doing', *gwiinawii* 'don't know, not able to', and *goji* 'try, attempt'. Still others pertain to the way that a speaker perceives an event: *bi* 'here, toward the speaker', *madwe* 'audible from a distance, can be heard', *ani/ni* 'going away from the speaker, approaching the time of doing something'. In short, preverbs have different functions that parallel phrasal expressions in nonpolysynthetic languages. Although it is possible for several preverbs to appear in a string, their order is not random.

One of the characteristics of a polysynthetic language is the richness of word formation processes. In Ojibwe/Algonquian, many words are made up of multiple roots, not unlike English compounds, e.g. 'killdeer' (a kind of bird), 'snow removal', etc. Roots can be divided into initials, medials, and finals, depending on their position within the word. Initials carry most of the meaning, often describing processes or resulting states. Medials are usually nounlike, denoting body parts or instruments, and finals typically indicate simple states or actions. The complex Passamaquoddy stem *napici koli-hqeh-mon* 'She or he sticks it onto something' could be translated as 'onto stuck-surface-make' [preverb initial-medial-final]. Final roots in Ojibwe/Algonquian also specify the gender class of intransitive subjects and transitive objects (cf. the discussion of gender below). A common pitfall in understanding polysynthetic languages has been to ascribe variations in word form to the speaker's viewpoint of the world, when in fact they probably result from regular grammatical processes.

The seemingly many names for snow in Eskimo (*Yup'ik, Inuktitut*) is a classic example. Similarly, Algonquian languages (like French) seem to contain an inordinate number of verbal paradigms. The difference, for example, between 'if he comes' and 'when he comes' depends on the change or deletion of a single vowel. The third major order in Algonquian is the imperative. The important thing to realize is that speakers do not memorize these myriad forms, but rather freely produce and understand them by internalizing a finite set of rules.

Nouns in Ojibwe (Algonquian) are classified on the basis of animacy, roughly 'being like a conscious living thing'. As one might expect, humans and animals are [+animate], whereas rocks and similar types of objects are [−animate]. The division is not always based on clear-cut properties, however. For example, trees are [+animate], along with containers used for liquid (cups, spoons, pens, etc.). In Passamaquoddy (Eastern Algonquian), the words for rope, milk, and fallen snow are all [+animate], as are those for fingernail and knee. Other body parts (heart, tongue) are [−animate], however. The animacy of a given noun is overtly indicated by the verb-stem final, but only under certain grammatical conditions.

Ojibwe speakers have a complex system of referring to things or people in a sentence or discourse. English normally uses pronouns like 'he' or phrases like 'that guy' to refer to someone after first introducing him by name. In Ojibwe, however, pronouns are not expressed overtly: one usually infers their presence through subject/object agreement on the verb or possessor agreement on the noun. The 'obviation' system guarantees that certain noun phrases (inaudible pronouns as well as audible noun phrases) are not confused with others. Within a complex noun phrase like *niw wday-an* 'his dog', for example, the possessive pronoun 'his' (which is only implied) would not be indicated by inflection, whereas *niw* 'dog' is marked as obviative (the suffix -*an*). This ensures that 'his' and 'dog' do not refer to the same entity. The second domain of obviation is the clause. This can be seen in *uki:-necci:we'a:-n [eniw kwi:wesse:ns-an]* 'He scolded [that boy]', where the subject is unmarked and the object triggers obviation (underlined) on the verb as well as the noun phrase. As in possessed noun phrases, obviation within a clause is obligatory. In the third domain of obviation, subjects of main clauses are unmarked, but those in subordinate clauses are obviative (underlined):

gii-boonii-w dash maa dVdibew [mitigoonsikaa-<u>ini</u>-g]
'Then she landed on the shore [where there were bushes]'

Further restrictions determine whether subjects and objects are overtly expressed or implied. Although it is possible to say 'I wrote them' or 'You hurt me', for

instance, the opposite 'They wrote me' or 'I hurt you' cannot be expressed without changing the affixes of the verb. In the first two sentences, a 'direct' morpheme is attached to the verb stem, and in the latter two an 'inverse' one. The choice of direct or inverse is determined by a 'participant hierarchy' that ranks noun phrases according to person and obviation. In direct sentences, subjects outrank objects on the scale: 2 > 1 > 3 > 3′—that is, second persons ('you') are ranked higher than first persons ('I/we'), which in turn are higher than third persons ('she or he'). Nonobviative third persons (3) are ranked higher than obviative (3′) ones. In inverse sentences, exactly the opposite holds: objects must outrank subjects on the same scale. Strikingly, subjects in the inverse appear as objects in the direct, and vice versa:

Direct	Inverse
ni-wa:pam-a:-k	ni-wa:pam-ik-o:k
1-see-third person/direct-pl	1-see-third person/inverse-pl
'I see them'	'They see me'

Hierarchical effects such as those exhibited by Ojibwe/Algonquian pose a challenge to linguists attempting to explain them.

References

Aanii: an introduction to the central Ojibwe language. Video Series. Lessons in conversational Ojibwe as spoken at Wikwemikong, Ontario. Videotapes with Ojibwe and English subtitles, pauses for repetition, and additional vocabulary. Native Vision Productions, Box 23, Ohsweken, Ontario N0A 1M0, Canada (tel.: 519/445-0639; fax: 519/445-0639; e-mail: mwwwturner@yahoo.ca; website: www.nativevision.homestead.com)

Baker, Mark. 1996. The polysynthesis parameter. Oxford: Oxford University Press.

Bloomfield, Leonard. 1946. Algonquian. Linguistic structures of native America, ed. by Hoijer Harry, 85–129. Viking Fund Publications in Anthropology, Vol. 6. New York: Viking.

———.1957. Eastern Ojibwa. Ann Arbor: University of Michigan Press.

Campana, Mark. 1996. The conjunct order in Algonquian. Canadian Journal of Linguistics 41(3). 201–34.

*Campbell, Lyle. 1997. American Indian languages: the historical linguistics of Native America. New York: Oxford University Press.

*Dahlstrom, Amy. 1991. Plains Cree morphosyntax. New York: Garland.

Goddard, Ives. 1978. Central Algonquian languages. Handbook of North American Indians 15. 583–7.

———.1990. Primary and secondary stem derivation in Algonquian. International Journal of American Linguistics 56(4). 449–83.

———. 1994. The east-west cline in Algonquian dialectology. Papers of the 25th Algonquian Conference, ed. by W. Cowan, Carleton University, 187–211. Ottawa.

Hockett, Charles. 1966. What Algonquian is really like. International Journal of American Linguistics 32. 59–73.

Leavitt, Robert. 1992. Lexical exploration and educational insight. Papers of the 23rd Algonquian Conference, ed. by W. Cowan. Carleton University, Ottawa.

———. 1996. Passmaquoddy-Maliseet. Languages of the world/materials, Vol. 27. Newcastle: Lincom Europa.

Lyovin, Anatole. 1997. An introduction to the languages of the world. Oxford: Oxford University Press.

Mithun, Marianne. 1999. The languages of native North America. Cambridge: Cambridge University Press.

Nichols, John. 1980. Ojibwe morphology. Doctoral Dissertation, Harvard University.

*Rhodes, Richard. 1985. Eastern Ojibwa-Chippewa-Ottawa dictionary. Berlin: Mouton de Gruyter.

———. 1990. Obviation, inversion, and topic rank in Ojibwa. Berkeley Linguistic Society 16. 101–15.

Siebert, Frank. 1967. The original home of the proto-Algonquian people. ed. by A.D. DeBlois, Contributions to anthropology: linguistics I (Algonquian), 13–47. Anthropological Series 78, National Museum of Canada Bulletin 214. www.ncs4.net/Ojibwe. The definitive link to Ojibwe linguistic and anthropological material. www.citilink.com. Nancy Vogt's website, with brief descriptions of language and culture; bibliography of further references.

MARK CAMPANA

Okanagan and Salishan Languages

Okanagan is one of four languages that belong to the Southern Interior branch of the Salishan linguistic family of North America. All but two languages of the family occupy contiguous territory that extends longitudinally from about 123°W to about 113°W, and latitudinally from about 52°N to about 45°N. In contrast to early speculations that the Salishan languages had spread from an inland location (Boas 1905), recently scholars have proposed that speakers of the original language occupied a maritime area, and migrated outward from there, following routes to the interior along such major rivers as the Fraser and Thompson (Suttles

and Elmendorf 1963; Suttles 1987). The linguistic evidence adduced focuses on terms of flora and fauna (Kinkade 1991).

Subgroupings: The family divides into five branches, the result of migrations. A single group is thought to have headed northwest and settled in the area where Bella Coola is now found, surrounded by non-Salishan languages, Wakashan seaside, and Athapascan inlandside. Another group is thought to have settled along the coast, and then spread further, one subgroup going southward to where Tillamook, now extinct, was spoken, and formed the third branch of the linguistic family. This language, too, is surrounded by non-Salishan languages, clockwise from the north: Chinookan, Athapascan, Takelman, and Maidu. A fourth group, forming the Tsamosan branch, also moved southward, and a fifth group, forming the Interior branch, moved eastward (see Figure 1). One now extinct Athapascan language, Nicola, was spoken in the approximate geographic center of the Salish area.

The area where Okanagan is spoken by perhaps as many as 1,000 speakers of mostly mature age spans along the north–south expanse of the Okanagan valley from what is now Enderby to the south of Okanogan,

Washington, and westward in the Similkameen and Methow valleys; and along the north–south expanses of the Sanpoil and Kettle rivers, and the area west of the Columbia river as far as the bend around Wilbur, Washington. Dialectal differences are minor but remain to be described. (Figure 1 here)

Phonology: In spite of the size of the family, and the number of different languages represented, it is possible to provide a generalized phonology of Salish (see Figure 2).

The angled brackets notation should be read as either/or: with few exceptions, a Salishan language has one or the other series, the velar being the conservative one. Okanagan has the velar series. The northern dialects of Okanagan include voiced postpalatal resonants (ɣ', ɣ) that correspond to y' y in the other dialects. In the same northern dialects, the labialized pharyngeals have merged with their unrounded counterparts. Voiced stops are rare in Salishan languages, but occur in Coeur d'Alene, Twana, Lushootseed, and Comox, where they do not derive historically from proto-stops. Pharyngeals are found only in the Salish languages of the Interior; and no language, except Comox, has a phonemic nonejective lateral affricate.
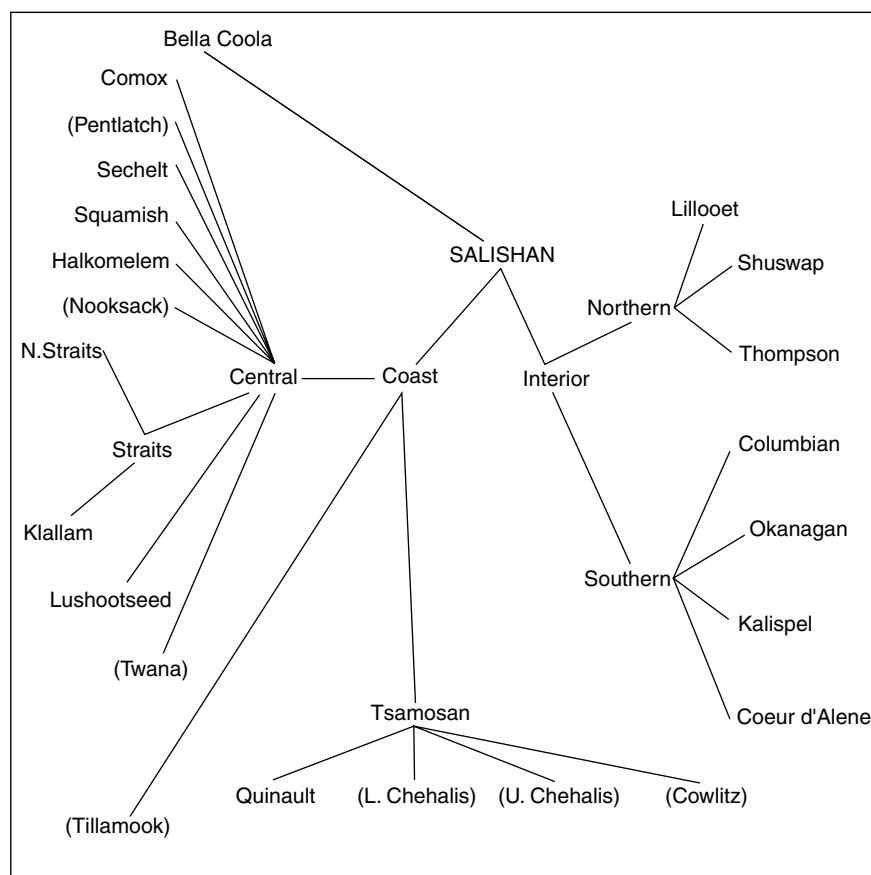
Figure 1

p'	t'	c'	λ'	<tʃ'>	<k'>	kʼʷ	q'	qʼʷ	
p	t	c		<tʃ>	<k>	kʷ	q	qʷ	ʔ
	s	ɬ		<ʃ>	<x>	xʷ	χ	χʷ	h
m'	n'	r'	l'		y'		w'	ʕ'	ʕʼʷ
m	n	r	l		y		w	ʕ	ʕʷ

Figure 2. Generalized Salish consonant inventory.

Okanagan has a symmetric three -vowel system, /i u a/, while the vowel systems of Salishan languages often include four vowels, sometimes five, with some asymmetries, notably the occasional absence of a high back vowel. Schwas are for the most part epenthetic. Salishan languages are wellknown for their propensity for consonant clusters; Bella Coola for several lexical items that consist entirely of voiceless obstruents.

Morphosyntax: Salish languages have large numbers of affixes, predominantly suffixes. All Salish languages have developed lexical affixes, bound forms with lexical content, with a function that resembles that of incorporated nouns, and all Salish languages have several reduplicative patterns that mark primarily augmentative, diminutive, and inchoative forms. A nominalizer with the shape *s-* is nearly universally found, but homophonous aspectual prefixes also occur, notably in Coeur d'Alene and Okanagan.

Person marking: Six Interior languages have different sets of person markers for transitive, intransitive, and subordinate predicates; the paradigms are less differentiated in Lillooet, closest to the languages of the coast, and in the noninterior languages. Possessive paradigms include prefixes and suffixes in all the languages. Okanagan has four main sets of person reference markers: the kn̲ set (intransitive), the i(n)- set (possessive), the -(í)n set (transitive subject), and the (transitive) object set.

The kn̲ set consists of clitics (marked with the ligature) and a suffix:

kn̲	1sg	kʷu̲	1pl
kʷ̲	2sg	p	2pl
Ø	3sg	Ø ...-lx	3pl

These markers accompany stems that in English translate as intransitive verbs, nouns, and adjectives.

kn̲ ʔitx. *I slept.*
kʷ̲sqilxʷ. *You are an Indian/a person.*
ʔayχʷt (axáʔ). *This one is tired.*

A subset of these markers, identical in all persons except for 1sg kʷu̲, co-occurs with the possessive set of person markers, and is reserved for double possessives and verb nominalizations.

The possessive set, used with nouns, psych verbs, and verb nominalizations, consists of these markers

(prefixes and suffixes; parentheses abbreviate variants):

i(n)-	1sg	-tt	1pl
a(n)-	2sg	-mp	2pl
-s/-c	3sg	-s-lx / -c-lx	3pl

which yield such forms as

an-lʔíw *your father*
in-χmínk *I like/want it*

which, in turn, may combine with members of the kn̲ set (kʷu̲ subset) to yield forms such as

kʷu̲ an-lʔíw *I am your father.*
kʷ̲ in-χmínk *I like/want you (you are my wanting).*
kʷ̲ i-ks-ʔam-ɬt-ím an-lʔíw *I am going to feed your father.*

The last is the nominalization of a future (ks-) possessor applicative (-ɬt) verb form (root ʔam, *feed*), in which the suffix -(i)m, sometimes referred to as the antipassive, is required.

The transitive subject set, often called the ergative set, consists of the following suffixes (parentheses abbreviate stressed and unstressed variants):

-(í)n	1sg	-(í)m /-t	1pl
-(í)xʷ	2sg	-(í)p	2pl
-(í)s	3sg	-(í)s-lx	3pl

These markers follow the object markers, which, in turn, follow one of several obligatory transitive markers (see below).

The (transitive) object set consists of the following markers (one proclitic and suffixes):

kʷu̲	1sg	kʷu̲...-m	1pl
-s / -m	2sg	-ɬ(úl)m	2pl
-Ø	3sg	-Ø... -lx	3pl

Because third -person object markers and thirdperson intransitive subject markers are Ø, Salishan languages are often characterized as split ergative systems. The allomorphy of the second singular object is transitivizer dependent. The disambiguation of number in the first-person object is accomplished by the suffix -m and such forms are interpreted as 3rd indef subject - 1pl object:

kʷu̲ sp'-nt-is *He whipped me (-nt transitivizer).*
kʷu̲ sp'-nt-im *They whipped us/We were whipped.*

-(í)m occurs also with Ø, and the interpretation of these forms can be indefinite subject, or passive:

| sp'-nt-is | *3rd person whipped 3rd person.* |
| sp'-nt-im | *3rd person indef whipped 3rd person/3rd person was whipped.* |

Word classes: Aspectual criteria can be established to distinguish word classes, and, as expected, these may derive forms of other classes—nouns can derive verbs and verbs can be nominalized (for example, N. Mattina as reported in Kroeber 1999). A prototypical noun like **kʷilstn** *sweat lodge*, culturally relevant and categorially marked (**-tn** instrumental), derives a verb with **-m**:

kn̓ kʷilstn-m. *I sweat bathed.*

Similarly, **qʷacqn** *hat* derives **qʷacqn-m** *wear a hat* (intransitive); **ntχʷχʷin** *noon* derives **ntχʷχʷin-m** *do lunch* (intransitive).

Analogously, **qiʔs** *dream* (intransitive) derives **s-qiʔs** *dream*, and the latter form can be inflected with possessive markers and interpreted as a possessive noun form, or as a nominalized verb form.

Most Okanagan stems can also be transitivized (see below).

Nominal and pronominal arguments: Scholars have argued that Salishan languages are pronominal argument languages: a form like **wik-nt-xʷ** *You saw it* is a full sentence with a third-person object (Ø) and second-person subject (-xʷ). In this interpretation, any object expressed in nominal form is an adjunct, not a (nominal) argument. The claim is countered with the suggestion that in applicative sentences like

kʷu̓tq-ɬt-is in-kílx. *He touched my hand*

the noun phrase in-kílx *my hand* functions as one of the arguments of the possessor applicative verb form kʷu̓tq-ɬt-is *He touched my ...* and this argument is not, and cannot be, referenced in pronominal form on the verb.

Intransitive forms are also analyzed as fully predicative.

kn̓ xʷuy	*I went.*
kʷ̓ilmíxʷm	*You are a boss.*
kʷ̓ χast	*You are fine.*

In these sentences, the clitics kn̓ and kʷ̓ are the subjects, and the word to which the clitics are attached are the predicates. Third-person forms have Ø subject person marking, and forms like **sql'tmixʷ** have been analyzed as full predications that should be translated as something like 'He is a man' or 'It's a man.' In the stream of discourse, such words can function as predicative elements. The normal way to express either of the isolated propositions 'He's a man' and 'It's a man' is with utterances like **ixíʔ sql'tmixʷ** *That's a man*, or

sql'tmixʷ yaʔχís *That one over there is a man*; that is, by juxtaposing (in either order) the stem **sql'tmixʷ** and a deictic stem (ixíʔ, yaʔχís). In traditional terms, these sentences would be analyzed as exocentric equational constructions consisting of a subject and a predicate. The participant persons kn̓ and kʷ̓ are pronominal subjects; third-person forms can be analyzed as having a nominal subject of the classes mentioned, which, in context, can be deleted. Another complication for the interpretation of all full words as predicative is presented by the different markings for morphological and syntactic plurals: the morphological plural of **citxʷ** *house* is the reduplicated form **ct-citxʷ** *houses*, while the syntactic plural of the same form is **citxʷ-lx** (ixíʔ) *(Those) are houses*.

In recent times, when scholars are preferring to view all constructions to have heads (or centers, in the old terminology), the question is raised as to what constitutes the head of such a sentence as kn̓ sql'tmíxʷ. Most common is the hypothesis that the verb is the head of the sentence (here it would be the predicate nominal), but because the identification of *head* with *lexical head* can be dispensed with, just as abstract features within the Inflection or Agreement nodes have been proposed to head sentences, and just as the determiner has been proposed to head Determiner Phrases, so can kn̓ be proposed to head the sentence kn̓ sql'tmíxʷ. An utterance like **xʷuy** *He went*, then, can be viewed as the abbreviation of **xʷuy ixíʔ** *That one went*, and analyzed either as having a null subject, or as requiring a third-person nominal subject which undergoes deletion in the appropriate circumstances.

Intransitive, possessive, and transitive paradigms: Beside the intransitive constructions already discussed, Okanagan uses kn̓ inflection in a number of forms derived by means of prefixes, suffixes, and circumfixes. Among these forms are: To-Be nouns (kn̓ kɬ-noun). **kɬ-ilmíxʷm** snk'lip *Coyote will be chief/is chief-to-be;* kɬ- *have* forms (kn̓ kɬ-noun). kn̓ **kɬ-qʷacqn** *I have a hat;* inchoatives (kn̓ verb+-ʔ- before stressed vowel). kn̓ **c'ʔax** *I got ashamed.* (root c'ax); patient forms (kn̓ verb+VC₂). kn̓ **t'k'ʷ-ak'ʷ** *I fell.* (cf. t'k'ʷ-nt *put something down*); get patient forms (kn̓ c+verb). uɬ ilíʔ kn̓ **c-lak'** *I was in jail* a long time (cf. lk'-nt *tie something*); habitual/durative forms (kn̓ c+verb). ilíʔ kn̓ **c-wix** *I live there;* imperfective forms (kn̓ s-c+verb-(mi)x *I have been X-ing*); inceptive forms (kn̓ ks+verb-(mí)xaʔx *I am about to X*); and past perfect forms (kn̓ ksc+verb). kn̓ ksc-nik' *I have cut some, I have some cut*.

Beside the possessive, double possessive, and psych forms discussed, other forms, some intransitive and others transitive (the latter always marked by the suffix -m), take possessive inflection: durative/intent forms (i+s+verb) s-q'sápiʔ-s ilíʔ **i-s-ilíʔ** *I lived there* a long

time (root ilíʔ *there*, lit. long-time there I-there); perfective forms (i+sc+verb). in-χást **i-sc-ʔítx** *I slept well* (my-good my-having-slept); future forms (i+ks+verb) lut a-**ks-xʷúy** *Don't go*; kʷ i-**ks-(s)íw-m** *I'll ask you*; future imperative forms (i+kc+verb) lut a-**kc-náqʼʷ** *You will not steal*. χast a-**kc-kʼʷúlʼ-m** *You will work well*; and future applicative forms (a-ks-verb-t-m) kw̰ i-**ks-mayʼ-xít-m** *... I am going to tell you ...*

Finally, all transitive forms take transitive person markers. Okanagan has two transitivizers, -nt and -st; a *causative* -st; three applicatives -ɬt, -x(í)t, -túɬt; and three suffixes that prepare stems for transitivization: -nun, -min, -xix. Customary transitive forms are marked with the circumfix c-...-st, as in c-wik-st-n *I always see it*.

References

Boas, Franz. 1905. The Jesup North Pacific expedition Proceedings of the 13th International Congress of Americanists, 91–100. New York, Easton, PA: Eschenbach Printing Co.

Czaykowska-Higgins, Ewa, and M. Dale Kindade (eds.) 1998. Salish languages and linguistics. Theoretical and descriptive perspectives. Berlin: Mouton De Gruyter.

Kinkade, M. Dale. 1991. Prehistory of the native languages of the northwest coast. Proceedings of the Great Ocean Conferences, Vol I: The North Pacific to 1600. Portland: Oregon Historical Society Press.

Kinkade, M. Dale, William W. Elmendorf, Bruce Rigsby, and Haruo Aoki. 1998. Languages. Plateau Handbook of North American Indians, Vol. 12, ed. by Deward E. Walker (William C. Sturtevant, general editor), 49–72. Washington, DC: Smithonian Institution.

Kroeber, Paul D. 1999. The Salish language family. Reconstructing syntax. Lincoln, NE: University of Nebraska Press.

Suttles, Wayne. 1987. Northwest coast linguistic history. A view from the coast. Coast Salish Essays, ed. by Wayne, suttles. Vancouver: Talonbooks.

Suttles, Wayne, and William W. Elmendorf. 1963. linguistic evidence for Salish prehistory. Proceedings of the 1962 Annual Spring Meeting of the American Ethnological Society, ed. by Viola E. Garfield, and Wallace L. Chafe. Seattle.

Thompson, Laurence C. 1973. The Northwest. Current trends in linguistics, Vol. 10. Linguistics in North America, ed. by Thomas A. Sebeok, 979–1045. The Hague: Mouton.

——— 1979 Salishan and the Northwest. The languages of Native North America: historical and comparative assessment, ed. by Lyle Campbell, and Marianne Mithun, 692–765. Austin: University of Texas Press.

Thompson, Laurence C., and M. Dale Kinkade. 1998. Languages. Northwest Coast of Handbook of North American Indians, Vol. 17, ed. by Wayne Suttles. (William C. Sturtevant, general editor), 30–51. Washington, DC: Smithonian Institution.

ANTHONY MATTINA

Old Chinese

Old Chinese may be subperiodized into Early Old Chinese, Middle Old Chinese, and Late Old Chinese. It is generally thought that most Early Old Chinese polysyllabic words became monosyllabic by or during the Middle Old Chinese period, so the Late Old Chinese language was overwhelmingly monosyllabic.

Proto-Chinese, the ancestor of Old Chinese, is practically unknown. Although many scholars believe it to be related to the Tibeto-Burman family of languages in a 'Sino-Tibetan' genetic family, the theory remains controversial due to the lack of regular correspondences in phonology (sound structure), morphology (word structure), and syntax (sentence structure) between Chinese and Tibeto-Burman. Moreover, from Early Old Chinese times to the present, the Chinese language has spread into territory inhabited by peoples who originally spoke other languages, and for centuries it has been spoken by far more people than any other language in the world. Without a thorough comparative study of loanwords in languages known to have bordered on the Chinese-speaking area in Antiquity, the affiliations of Chinese must be considered uncertain and the relationship with Tibeto-Burman likely to be due to convergence. Even if the Sino-Tibetan theory is correct, Proto-Chinese and the Tibeto-Burman languages must have diverged long before the Oracle Bone Inscriptions, the first documents written in Old Chinese. These texts appear in the mid-second millenium BCE and already contain distinctively Chinese phonological and syntactic characteristics that had developed in Proto-Chinese, before the language was first written down. For example, only one negative particle, *ma, is reconstructable for Proto-Tibeto-Burman, but the Oracle Bone Inscriptions have two negative roots, *pa- and *ma-. This distinction is preserved in all later forms of Chinese, including modern Mandarin, where it is still found in numerous bound forms and in the free negative forms 不 *bù* [pu] and 沒 *méi* [mej].

The Old Chinese writing system contains important information for the reconstruction of the language.

Although it includes many characters with no phonetic elements, such as 龜 *guî* 'tortoise, turtle' (originally a pictograph), some characters are phonetically 'borrowed'. (N.B.: Pronunciation of characters is given in Mandarin unless otherwise noted.) For example, the originally pictographic character used to write the word for 'wheat' was borrowed to write the then homonymous word 來 *lái* 'to come'. Most characters are actually constructed of two or more parts, of which one part is a 'phonetic' element and another is an often simplified semantic or 'significant' element usually called the 'radical', such as 隹 *zhuî* 'short-tailed bird' (originally a pictograph) and 萑 **zhuî* (now usually pronounced, irregularly, *huán*) 'grass used for making mats', written with the 'grass' radical as the significant and 隹 *zhuî* as the phonetic. Although other examples—such as 魯 *lǔ* 'blunt, stupid; Lu, the home state of Confucius' and its phonetic 魚 *yú* 'fish'—have differences that are important for reconstruction (see below), the phonetic information contained in the script is not by itself sufficient to allow reconstruction of Old Chinese phonological structure. In fact, Chinese is not well attested phonetically until the seventh century CE, when Middle Chinese forms begin to be recorded in Old Tibetan alphabetic script. The phonology of Old Chinese can therefore only be recovered through reconstruction.

Modern reconstructions of Old Chinese nearly all belong to a tradition that may be called Historic Sinological Reconstruction. It derives ultimately from the early Chinese grammarians, on whose work the pioneering Sinologist Bernhard Karlgren based his system of reconstruction. Although the appearance of most Old Chinese reconstructions has changed radically since the publication of Karlgren's etymological dictionary *Grammata Serica recensa* (1957; A Chinese Grammar, Revised), Historic Sinological Reconstruction is still based on the method pioneered by Karlgren.

The most important sources used in the traditional Historic Sinological Reconstruction approach to Old Chinese phonology are: the phonological information derivable from analysis of the characters themselves (mostly mid-second millenium BCE to second century BCE) and their variants or substitutions; the rhymes of the *Shijing* (late first millenium BCE; Book of Songs); Chinese transcriptions of known foreign words (c. first century CE onward); books by Chinese grammarians on dialects and rhymes (first century BCE onward), especially the lost Middle Chinese *Qieyun* (601 CE; Cut Rhymes); the rhyme books, particularly the *Guangyun* (Extensive Rhymes) and *Jiyun* (Collected Rhymes), both from the eleventh century CE; and, above all, the rhyme tables, especially the *Yunjing* (Mirror of Rhymes), from the twelfth century CE. The latter three works, although compiled much later, are based ulti-

mately on the *Qieyun*. However, unlike earlier works, the *Yunjing* organizes the material in tabular form. Historic Sinological Reconstruction of Old Chinese depends heavily on the projection back in time of the categories established by the tables in the *Yunjing*, adjusted according to the rhyme categories implied by the *Shijing* poems and to phonological information derivable from the characters themselves. Change in one part of the system necessarily entails change in many other parts in order to maintain consistent correspondences between the categories of the tables and the phonological categories of Middle Chinese and the modern Chinese dialects, hence the reconstructions of one scholar's system often look radically different from those of another. Unfortunately, Historic Sinological Reconstruction does not allow subperiodization of Old Chinese—contemporary proponents of the method explicitly claim that they are not actually trying to reconstruct a real language spoken by real people in different places and at different times but only a theoretical construct from which all later forms of Chinese can be derived—so it is impossible to systematically use contemporaneous data to check the reconstructions themselves. This results in a tendency to ignore attested data in favor of the system. For example, Sergei Starostin (1989) reconstructs 女 'woman; you (in the latter sense also written 汝)' as Old Chinese **nraʔ* even though overwhelming Oracle Bone Inscription and other evidence indicates that the word initial must have been a bilabial in Early Old Chinese and apparently well into the Middle Old Chinese period. Another result is the production of many unlikely forms, such as Old Chinese **fisrjɨʔ(s)*, the reconstruction of 事 'affair, matter' by William Baxter (1992). Although scholars working in this tradition have made progress toward the reconstruction of Old Chinese, recently their method has been challenged. Reconstructions have been proposed that are not based on the categories of the rhyme tables but on contemporaneous data, emphasizing Old Chinese character variations, loanwords found in neighboring languages, and Middle Chinese reconstructions grounded in foreign transcriptions and the earliest attested text of the *Qieyun* (706 CE).

Early Old Chinese (called by some scholars 'pre-Archaic Chinese'), the language of the Yin, or late Shang dynasty period (fourteenth to eleventh centuries BCE), is preserved mainly in the Oracle Bone Inscriptions, which are divination texts inscribed on bones and turtle shells. These texts have been found in great numbers in the area of the lower Yellow River valley. The syntax of Early Old Chinese, although slightly different from that of the earliest classical texts from a millenium later in the late Middle Old Chinese and Late Old Chinese periods, is already SVO (Subject–Verb–Object) in basic sentence word-order,

as in modern Chinese. Several scholars have argued that Early Old Chinese must have had many disyllabic words or roots. Reconstructable examples include 女 Early Old Chinese *Cwena (where 'C' stands for an unknown consonant) 'woman; you', and 苦 Early Old Chinese *kara 'bitter'. However, Early Old Chinese phonology has yet to be studied in depth.

Middle Old Chinese is essentially the language of the post-Shang bronze and stone inscriptions and the earliest transmitted Classical texts. The inscriptions on bronze vessels are often dated, and cover the period from the early Zhou dynasty (c. 1145 BCE onward) into the Warring States period (475–221 BCE). Some of them contain rhymed passages and variant characters from which phonological information can be derived. The Middle Old Chinese period may be divided into three subperiods: the early Middle Old Chinese of the Western Zhou dynasty, when the capital was located in the area of present-day Shaanxi Province and the prestige language was influenced by the local dialect; the middle Middle Old Chinese of the Eastern Zhou, when the capital moved eastward into another dialect area; and late Middle Old Chinese, the language of the earliest transmitted 'Confucian' classical texts, including the *Shijing*. These periods are marked by phonological changes, partly motivated by prestige-dialect shift when the capital moved to a new dialect region, and partly due to change over time. One example must suffice here. The usual first person singular pronoun ('I') in early Middle Old Chinese is 余 (also written 予) *yú,* but in middle Middle Old Chinese (beginning with the Stone Drum inscriptions, c. fifth to sixth centuries BCE) this word is replaced by 吾 *wú,* which by Late Old Chinese displaces 余 *yú* except in deliberately archaizing texts. The character 吾 *wú* 'I' has as its phonetic element the character 五 *wǔ* 'five', while in other middle Middle Old Chinese texts it is generally written instead with 魚 *yú* 'fish' as its phonetic. Because the Old Tibetan numerals are known to be cognate with (derived from the same historical source as) Chinese, Old Tibetan *lŋa* 'five' is cognate with the Chinese word for 'five'. (Whether the Tibetan numerals are borrowed from Chinese or inherited from a common 'Sino-Tibetan' ancestor is irrelevant for present purposes.) Since 魚 *yú* (from Middle Chinese *ŋɨɔ[1]) 'fish' is also the phonetic in the character 魯 *lǔ* (from Middle Chinese *lɔ[2]) 'Lu', it is clear that not only must the Middle Old Chinese form of the word for 'five' have been close to *lŋâ, but it and the word for 'fish', the first person pronoun, 'Lu', and other words, such as 語 *yǔ* 'speak, speech', were pronounced the same in one or more of the Middle Old Chinese dialects. Because the character used to write the earlier first person pronoun 余 *yú* is also used as a phonetic in other characters, it is reconstructable as *lâ, which derives from an earlier *laCa. It is clear that these

synonyms must be related. Since 吾 *lŋâ, one dialect form of the Middle Old Chinese first person pronoun, may be reconstructed as *laŋa, the other dialect form, 余 *lâ, may be reconstructed as *laɣa, both deriving from an Early Old Chinese *laga or *lege. This word is evidently related in turn to another early first person pronoun, 我 Early Old Chinese *aga or *ege.

During the Middle Old Chinese period, most previously disyllabic morphemes became monosyllabic. This change produced many homonyms by late Middle Old Chinese—such as *lŋâ 'first person pronoun' ~ 'five' ~ 'fish' ~ 'speak'—and resulted in other drastic phonological changes, including widespread metathesis (segments changing place) in one or more dialects. For example, middle Middle Old Chinese *lŋâ 'fish' (from Proto-Chinese *laka) became late Middle Old Chinese *ŋlâ or *ŋrâ. The development of monosyllabism was also probably a major motivation for the eventual development of phonemic tone (intonational pitch distinguishing meanings). The shift to a monosyllabic morpheme structure was largely complete by the time the *Shijing* was recorded.

Late Old Chinese is characterized by a number of phonological changes, including that of the Middle Old Chinese syllable-initial *l to either *y [j] or *d, of Middle Old Chinese medial *l or *r to *ɨ, and of the Middle Old Chinese syllable-final *r to *n or *y [j]. (The rules for these changes are still unclear.) The change of final *r to *n took place in the ancient central dialect and is found in all modern Chinese dialects, but it was not completed in some peripheral dialects, notably in the northeast, where ancient final *r was retained as a syllable coda at least into the Middle Chinese period. Another significant change includes the beginning of what has been called the Great Chinese Vowel Shift, in which the Middle Old Chinese vowel *â (or [aː]) became Middle Chinese *ɔ and Mandarin *u*. Thus, late Middle Old Chinese *ŋlâ (or *ŋrâ) 'fish' became Late Old Chinese *ŋɨâ, Middle Chinese *ŋɨɔ[1], Late Middle Chinese *ŋü[1], and Mandarin *yú* [ü]. Some traces of Late Old Chinese forms are found in the modern Southern Min dialect. (Chinese is otherwise internally reconstructable only back to Middle Chinese.) The written Late Old Chinese of the Former Han dynasty (206 BCE to 9 CE) became the standard Classical Chinese literary language, which continued in use into the twentieth century and is still used for some purposes even today.

References

Baxter, William H. 1992. A handbook of Old Chinese phonology. Berlin: Mouton de Gruyter.

Beckwith, Christopher I. 2002. The Sino-Tibetan problem. Medieval Tibeto-Burman Languages, ed. by Christopher I. Beckwith. Leiden: E. J. Brill.

Beckwith, Christopher I. 1999. Review of Laurent Sagart, The roots of Old Chinese. Amsterdam: John Benjamins; in Anthropological Linguistics 44.2 (2002).

Benedict, Paul. 1972. Sino-Tibetan: a conspectus. Cambridge: Cambridge University Press.

Karlgren, Bernhard. 1957. Grammata Serica recensa. Stockholm: Museum of Far Eastern Antiquities; reprinted 1972.

Pulleyblank, Edwin G. 1962. The consonantal system of Old Chinese. Asia Major 9.

———. 1996. Early contacts between Indo-Europeans and Chinese. International Review of Chinese Linguistics 1.1.

Sagart, Laurent. 1999. The roots of Old Chinese. Amsterdam: John Benjamins.

Schuessler, Axel. 1987. A dictionary of early Zhou Chinese. Honolulu: University of Hawaii Press.

Starostin, Sergei A. 1989. Rekonstrukcija drevnekitajskoj fonologičeskoj sistemy (Reconstruction of the Old Chinese phonological system). Moscow: Nauka.

CHRISTOPHER I. BECKWITH

Old Church Slavonic

Old Church Slav(on)ic (OCS) is the language of a group of Slavic texts containing copies of religious works, almost all originally translated from Greek. The corpus includes eight parchment manuscripts having 100 or more folia, and an approximately equal number of fragments. A half-dozen inscriptions survive from the same period and area. No manuscript bears an explicit date or locale, but paleographic and linguistic features indicate that most of the surviving manuscripts were copied in the eleventh century CE, with a very small number possibly from the late tenth. The original translations are generally assumed to go back to the second half of the ninth or early tenth centuries, and were connected with the mission of Cyril and Methodius to the Slavs and the activities of their pupils in Bulgaria following the death of Methodius in 885 CE. Every manuscript deviates in some ways from normalized Old Church Slavonic, but as a group they show an Eastern South Slavic dialect which has often been called Old Bulgarian, and it is in fact very close to the reconstructed Late Common Slavic (LCS), which is usually considered to have lasted until around the end of the first millennium CE. With the passage of time, this liturgical and literary language took on certain local characteristics in each area where it remained in use, producing Russian/Bulgarian/ Serbian, etc. recensions of Church Slav(on)ic, which remained mutually comprehensible and collectively served as a common literary language for the Orthodox Slavs and some other peoples, such as the Romanians.

Old Church Slavonic was written in two alphabets: Glagolitic, invented (or, less likely, edited and brought into its final form) by St. Cyril, perhaps with the participation of his brother Methodius, for their mission to the Slavs in the 860s, and Cyrillic, invented later, probably in Eastern Bulgaria, and named in honor of Cyril. Cyrillic was based upon the Greek uncial letters; Glagolitic has no immediately obvious resemblance to any other alphabet, although some features compare suggestively to the Greek cursive of the time. Glagolitic was used in some parts of Croatia until the early twentieth century, but elsewhere (Kievan Rus', Bulgaria, Serbia, etc.) was quickly displaced by Cyrillic, which has remained in use among the Orthodox Slavs until the present day.

Late Common Slavic introduced far-reaching changes to the sound system of late Indo-European (IE). As in some other dialects of IE, *o and *a conflated to *å, with the long vowel giving Slavic a and the short vowel giving Slavic o. IE *ē became LCS *ě (in OCS, a low front vowel [æ]), and IE *ĕ went to LCS *e. IE *ī gave LCS *i, IE *ĭ gave LCS *ь, a high short (lax) front vowel. IE *ū eventually became *y, a high unrounded nonfront vowel (*sūnus > *synъ 'son'), and IE *ŭ became LCS *ъ, a high short (lax) back vowel.

This gave a vowel system with binary oppositions in height, fronting, and length (long/short):

	front	nonfront
high	ī / ĭ	ū / ŭ
nonhigh	ē / ĕ	ā / ă

realized as OCS

	front	nonfront
high	i / ь	y / ъ
nonhigh	ě / e	a / o

plus diphthongs. Those diphthongs in *w and *y were converted to monophthongs: *ow > *u; *ew > *ju; *oy > *ě or *i; *ey > *i. Diphthongs in *r and *l were transposed (*gord- > *grad- 'enclosed area,' *xold- > xlad- 'cold', and diphthongs in *m or *n became nasal vowels (*ronka > *rǫka 'hand,' *sēmen > *sěmę 'seed').

As a result, the vowel system of OCS contained i, ь, ě, e, a, o, y, ъ, u, ę, ǫ, with oppositions based upon front(i, ь, ě, e, ę)/back, high(i, ь, y, ъ, u)/low, rounded(ъ,u, o, ǫ)/unrounded, tense/lax, and oral/nasal.

Slavic is a *satem* language, as shown by such OCS words as *sъto* 'hundred' and *zlato* 'gold.' The consonant system shows only a two-way contrast, between voiced and voiceless, with a division into labial, dental, palatal (*not* palatalized), and velar. Stops were *p b t d k g*, continuants included *s z š ž x*, and the affricates were *c (ts)*, ʒ *(dz)*, č, and in some dialects, ǯ. Sonorants were *m n l r* ń *(=nj)*, ļ *(=lj)*, ŗ *(=rj)*, glides contained *w* and *j*.

LCS syllable structure had undergone a number of significant changes. Syllables in LCS were always 'open' Syllable-closing consonants were dropped: *sup-nos* > *sъnъ* (*sŭnŭ*). The basic syllable formula was *strV*, where *s* stands for *s* or *z*, *t* for almost any consonant, and *r* for a sonorant, although the details are quite complex. The consonant clusters that actually occur can be quite formidable: OCS *umrъštvļǫ* 'I will kill.'

Syllables also showed a type of 'synharmony', where consonants and vowels were subject to mutual accommodation, such as fronting of vowels after *j* or palatal consonants, the 'palatalization' of the velars *k g x* to *č ž š* or *c ʒ s* before a front vowel, or 'jotization' of combinations of consonants with *j*: *sj > š, kj > č*, etc. This produced paired endings and consonant alternations in several instances. For examples of the changes, see the *o/jo* and *a/ja* declensions given below, including the forms of *vlьkъ* and *rǫka* given in the same tables.

TABLE 1

(a) *Singular*

Gloss	stem	nom.	voc.	acc.	gen.	loc.	dat.	instr.
'city'	*o*	gradъ	grade	gradъ	grada	gradě	gradu	gradomь
'man'	*jo*	mǫžь	mǫžu	mǫža	mǫža	mǫži	mǫžu	mǫžemь
'wolf'	*o*	vlьkъ	vlьče	vlьkъ	vlьka	vlьcě	vlьku	vlьkomь
'village'	*o*	selo	selo	selo	sela	selě	selu	selomь
'field'	*jo*	poļe	poļe	poļe	poļa	poļi	poļu	poļemь
'woman'	*a*	žena	ženo	ženǫ	ženy	ženě	ženě	ženojǫ
'soul'	*ja*	duša	duše	dušǫ	dušę	duši	duši	dušejǫ
'hand'	*a*	rǫka	rǫko	rǫkǫ	rǫky	rǫcě	rǫcě	rǫkojǫ
'bone'	*ĭ*	kostь	kosti	kostь	kosti	kosti	kosti	kostьjǫ
'son'	*ŭ*	synъ	synu	synъ/-a	synu	synu	synovi	synъmь

(b) *Plural*

meaning	nom./voc.	acc.	gen.	loc.	dat.	instr.
'city'	gradi	grady	gradъ	graděxъ	gradomъ	grady
'man'	mǫži	mǫžę	mǫžь	mǫžixъ	mǫžemъ	mǫži
'wolf'	vlьci	vlьky	vlьkъ	vlьcěxъ	vlьkomъ	vlьky
'village'	sela	sela	selъ	selěxъ	selomъ	sely
'field'	poļa	poļa	poļь	poļixъ	poļemъ	poļi
'woman'	ženy	ženy	ženъ	ženaxъ	ženamъ	ženami
'soul'	dušę	dušę	dušь	dušaxъ	dušamъ	dušami
'hand'	rǫky	rǫky	rǫkъ	rǫkaxъ	rǫkamъ	rǫkami
'bone'	kosti	kosti	kostьjь	kostьxъ	kostьmъ	kostьmi
'son'	synove	syny	synovъ	synъxъ	synъmъ	synъmi

(c) *Dual*

meaning	nom./voc./acc.	gen./loc.	dat./instr.
'city'	grada	gradu	gradoma
'man'	mǫža	mǫžu	mǫžema
'wolf'	vlьka	vlьku	vlьkoma
'village'	selě	selu	seloma
'field'	poļi	poļu	poļema
'woman'	ženě	ženu	ženama
'soul'	duši	dušu	dušama
'hand'	r ǫcě	rǫku	rǫkama
'bone'	kosti	kostьju	kostьma
'son'	syny	synovu	synъma

TABLE 2

	present	root aorist	s-aorist	extended a.	imperfect	imperative
1 sg.	vedǫ	vedъ	věsъ	vedoxъ	veděaxъ	–
2 sg.	vedeši	vede	vede	vede	veděaše	vedi
3 sg.	vedetъ	vede	vede	vede	veděaše	vedi
1 dual	vedevě	vedově	věsově	vedoxově	veděaxově	veděvě
2 dual	vedeta	vedeta	věsta	vedosta	veděašeta	veděta
3 dual	vedete	vedete	věste	vedoste	veděašete	–
1 pl.	vedemъ	vedomъ	věsomъ	vedoxomъ	veděaxomъ	veděmъ
2 pl.	vedete	vedete	věste	vedoste	veděašete	veděte
3 pl.	vedǫtъ	vedǫ	věsę	vedošę	veděaxǫ	–

Inflected words in OCS may be divided into nominal and verbal groups, with the nominals subdivided into nouns, adjectives, and pronouns. Numerals pattern with nouns or pronouns; 'one' through 'four' and their compounds also show gender agreement.

Both verbs and nouns show number (singular, plural, and dual). Nouns distinguish up to seven grammatical cases marking subject (nominative), direct object (accusative), possession (genitive), location (locative), indirect object (dative), instruments/means (instrumental), and also a form of address(*vocative*). Nouns have inherent gender (masculine, feminine, and neuter). Verbs distinguish six tenses, and have indicative, imperative, and conditional forms; they can be active or passive and indicate completed vs. incompleted actions.

Noun markings in OCS fall into five patterns (declensions); in Indo-European terms they are *o/jo, a/ja,* ĭ, ŭ, and consonant stems. The difference between the *o* and *jo* stems and between the *a* and *ja* stems are due largely to the intrasyllabic changes referred to above as 'synharmony'. The *o* and the ŭ stem endings have intermixed so much by the time the OCS texts were copied that one can no longer really set up ŭ-stems as a separate declension. Consonant stems come in several varieties, e.g. *r*-stems such as *mati*, gen. sg. *matere* 'mother'; *s*-stems such as *nebo*, gen. sg. *nebese* 'heaven'; *en*-stems such as *vrěmę*, gen. sg. *vrěmene* 'time' and *kamy*, gen. sg. *kamene* 'stone'. The productive declensions are the *o/jo, a/ja,* and ĭ-stems. Some of the ŭ-stem endings occur frequently and even become productive in the *o/jo* declension.

Sample Declensions

See Table 1.

Adjectives are inflected and show agreement in gender, number, and case. Masculine and neuter adjectives decline in principle like *o/jo*-stem nouns, while feminine adjectival forms are essentially like those of the *a/ja*-stem nouns. Adjectives also distinguish between definite and indefinite*;* definite adjectives are made from indefinite adjectives by appending pronominal suffixes to the indefinite forms; thus, *nova rěka* means 'a new river,' but *novaja rěka* means 'the new river'; *novo selo* means 'a new village' and *novoje selo* (*o > e* after *j* by 'synharmony') means 'the new village.' The comparative (more...) is formed by suffixation; the superlative (most ...) is made syntactically from the comparative.

Although no single OCS verb is attested in all possible forms, the conjugation of *ved-* 'lead' must have been as shown in Table 2.

Word order in OCS is most often S(ubject) V(erb) O(bject), but can be quite free. It is often quite difficult to know when OCS word order is following that of the Greek from which the text was originally translated and when it represents truly Slavic usage. All grammatical cases except the nominative occur with prepositions, but all, including the locative, occur without prepositions. Under the influence of Greek, OCS syntax often becomes rather convoluted.

OCS vocabulary is primarily of Indo-European, or at least Balto-Slavic origin, although it also contains borrowings from several other sources, including Iranian (e.g. *bogъ* 'god,' *rajь* 'paradise') and Germanic (e.g. *xlěbъ* 'bread,' *kupiti* 'buy'). As one would expect, most of the religious terminology and an overwhelming majority of the abstract, legal, philosophical, administrative, and didactic vocabulary is either direct borrowings or loan translations (calques) from Greek.

References

Huntley, David. 1993. Old Church Slavonic. ed. by Bernard Comrie and Greville G. Corbett. The Slavonic Languages, London, New York: Routledge.

Lunt, Horace G. 2001. Old Church Slavonic Grammar, 7th revised edition. Berlin, New York: Mouton de Gruyter.

Schenker, Alexander M. 1995. The Dawn of Slavic. An Introduction to Slavic Philology. New Haven: Yale University Press.

CHARLES GRIBBLE

Old English

Old English (OE) is the collective name given to the varieties of Germanic brought to Britain by the Jutes, Saxons, and Angles in the fifth and sixth centuries. The event separating Old from Middle English is the Norman Conquest of 1066. The root *Engl-* in the names *Engl-ish* and *England < Engla-land* 'land of the Angles' represents a form of the Latin and Common Germanic name of the tribe *Angli*. The OE language is also sometimes referred to as Anglo-Saxon (AS), although more frequently a distinction is made between the OE language and AS history, law, literature, culture, etc. The closest linguistic relative of OE is Old Frisian. OE and Old Frisian belong to the West Germanic subgroup of the Germanic branch of Indo-European.

The date usually associated with the beginning of OE is 449 CE, when Germanic-speaking warriors, led by the legendary brothers Hengist and Horsa, sailed to Britain to fight against the Picts at the invitation of the Celtic king Vortigern. From the middle of the fifth century onward, Germanic settlers arrived in Britain in considerable numbers. The Jutes remained mostly in Kent, parts of Hampshire, and the Isle of Wight, and the Saxons occupied the lands south of the Thames, as well as Middlesex and Essex. The Angles spread westward and as far north as the Scottish Lowlands. The Germanic invasions and settlements resulted in the partial displacement of the indigenous Celtic-speaking populations from central England to the more southern, western, and northern parts of the country. The most common periodization of OE is into early OE, from the beginnings to c. 800 CE, classical OE, c. 800–950, and late OE, c. 950–1100.

Two notable historical events that had a profound effect on AS culture and the vocabulary of OE are the conversion to Christianity, initiated by the arrival of St. Augustine in Kent in 597, and the Viking invasions and settlements. The Vikings, who were Scandinavian seafaring warriors, began their raids on Northumbria in the late eighth century. In the ninth century, their attacks culminated in the establishment of a separate administrative unit, the Danelaw (< Dena lagu 'law of the Danes'), which included large parts of the northern, central, and eastern regions of England, roughly northeast of a line linking London and Chester. The treaty which established the Danelaw, negotiated by the West Saxon king Alfred the Great, who ruled from 871 to 889, provided for a relatively peaceful period during the tenth century. At the beginning of the eleventh century, the balance of political power shifted again toward Scandinavia: between 1016 and 1042, England was ruled by the Danish king Cnut and his sons.

Orthography

The earliest OE written records are runic inscriptions dated c. 650–700 CE. They were either records of individual names, or brief decorative messages. The runic alphabet, originally linked to Germanic pagan rituals, was abandoned after the adoption and spread of Christianity during the sixth and seventh centuries. The alphabet used subsequently was a modified form of the Roman alphabet, known as the *Insular hand* in which the letter <s> was written in its long form <ʃ>, and <g> appeared as <ȝ>, known as 'yogh'. The three specifically OE letters, not used in the Roman alphabet, are <æ> 'ash', for the vowel [æ], and <þ> 'thorn' and <ð> 'eth' (or 'ðæt'), used interchangeably for the consonant sounds [ð], as in *that*, or [θ], as in *thump*. Consonant letter combinations specific to OE writing are: <sc>, which represented [sk] for most of the period, but probably also [ʃ] after c. 1000, and <cȝ>, which stood for [ǰ].

Dialects

The three original groups of settlers, the Jutes, the Saxons, and the Angles, maintained their dialect differences throughout the AS period. The main dialects of Old English are Kentish, West Saxon, East Saxon, Mercian, and Northumbrian. Mercian and Northumbrian were two varieties of Anglian, spoken to the south and north of the river Humber. Prior to the Scandinavian invasions, Northumbrian was a dialect of great prestige due to the religious, artistic, and intellectual achievements of the Northumbrian kingdom. After the end of the ninth century, the kingdom of Wessex unified and dominated the rest of the country, and the new political and cultural situation led to the spread of the West Saxon literary norms to the neighboring dialects. West Saxon is the tenth century literary *koiné* of Anglo-Saxon England. Classical West Saxon is the variety of OE described in the standard reference works, and it is also the dialect on which the following descriptions are based.

Phonology

The system of OE stressed vowels consisted of seven short vowels, seven long vowels, and two diphthongs. The colon sign in parentheses is the phonetic symbol for vowel length. Vowel length was not marked in the

OE manuscripts, although many modern printings use the convention of a macron (ˉ) over a vowel letter to indicate that it is long.

Simple vowels   Diphthongs

i(:)   y(:)   u(:)   eo or eə
    e(:)   o(:)   æa or æə
        æ(:)   ɑ(:)

The most important phonological change, which occurred prior to the first OE written records and changed the quality of many vowels in some positions, was *I-Mutation*, also known as *I-Umlaut*. This is a right-to-left (regressive) vowel harmony phenomenon: back vowels became front and low front vowels were raised before /i, j/ in the same word. (The dashed line indicates the direction of the sound change from early OE to late OE and ME.)

i(:)<------ y(:) <——————   u(:)
↑
e(:) <------ ø(:) <——————   o(:)   . . . before /i,j/
↑
æ(:) <————————————   ɑ(:)
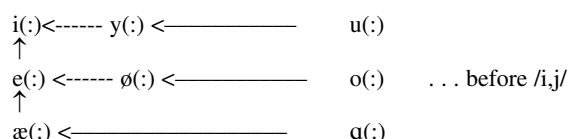
The change created stem alternations, depending on whether the stem was used by itself or whether it was followed by an affix containing /i,j/. I-Mutation is responsible for the different vowels in pairs such as *full-fill* (OE full - fyllan < *fulljan 'to fill'), *foot-feet* (OE fōt – fēt < *fōtiz 'feet', pl.), *man-men* (OE mann – menn < *manni- 'men', pl.).

An important quantitative change during the OE period was the lengthening of short vowels before some consonant groups, most notably *-ld, -mb, -nd*. This development, which started during the ninth century, accounts for the historically long vowels in words such as *child, comb, hound, kind*. A third consonant prevented the lengthening; this is the reason for the survival of the short vowels in words such as *children, hundred, kindred*.

Only short vowels could appear in fully unstressed syllables. The most frequent spellings of unstressed vowels are <e>, <o>, or <a>. In late OE, the spelling distinctions for unstressed vowels were disappearing, suggesting that by the end of the period most of the unstressed vowels were neutralized to [ə].

Classical OE had the following consonants:

	Labial	Dental	Alveolar	Palatal	Velar
Voiceless stops	p	t			k
Voiced stops	b	d			g
Fricatives	f/v	θ/ð	s/z	(ʃ)	x,ɣ
Affricates			(č),ǰ		
Nasals	m	n			
Liquids		l, r			
Approximants				j	w

The palatal fricative ʃ and the affricate č are in parentheses because their existence as independent units before the end of the OE period has not been proven. The voiced fricatives v, ð, and z could appear only word-medially when flanked by vowels or nasals and liquids. In all other positions, the fricatives were voiceless. The distribution of these sounds in OE is responsible for present-day alternations of the type *leaf-leaves* (OE *leaf*, sg.-*leafa* gen. pl, *leafum*, dat. pl. 'leaves'), *bath-bathe* (OE *bæþ*, n.-*bæþian* 'to bathe'), *glass-glaze* (OE *glæs*,n.-*glæsen*, adj. 'made of glass, vitreous, glazed'). All OE consonants except the approximants could be doubled in the word-medial position.

Word stress in OE was regularly placed on the first syllable of the root. This correlates with verse composition in OE, where alliteration on the first stressed syllable in each half-line is one of the basic structural characteristics of the line. The most frequent prefixes, *ge-* and *be-*, were never stressed, but some noun and adjective prefixes that retained their lexical meaning could be accented: *óferfyllo*, n. 'repletion' but *oferfýllan*, v. 'feed to excess', *andswárian*, v. 'answer', *ándswaru*, n. 'answer'. In compounds, both roots were stressed on the first syllable, with the first stress stronger than the second: *níed-fàru* 'enforced journey, death', *túngol-wìtega* 'astrologer', *scéarp-ècged* 'sharp edged'.

Morphosyntax

OE was a predominantly synthetic language, which means that the grammatical relations between the words in the clause were expressed by means of inflexional affixes. The nouns had special markings for the following categories:

(a) Number: singular, plural, e.g. *stan* 'stone', sg. *stan-**as*** 'stones', pl.

(b) Case: nominative (for the subject), accusative (for the direct object), genitive (noun modifiers, possessives), dative (indirect and prepositional objects), e.g. *stan-**es***, 'of stone, stone's', gen, *stan -**um*** 'with stones', dat. pl.

(c) Gender: masculine, feminine, neuter. The grammatical gender of a noun was determined by agreement with the demonstrative pronouns and adjectives accompanying it, not by reference to biological sex, e.g. *se stan* 'this stone', masculine, *þæt wif* 'this wife', neuter, and *seo rod* 'this cross', feminine.

The adjectives in OE had separate markings for the same categories as the nouns they modified (number, case, gender). They also varied according to their syntactic placement. The *strong,* or *definite* declension

was used for adjectives which were part of a definite noun group (*se brada sæ* 'this broad sea'), and the *weak*, or *indefinite* declension was selected when the adjective was used predicatively (*wæs þæt rice brad* 'that kingdom was broad'). Like modern adjectives, OE adjectives could be marked for degrees of comparison: comparative (*brad -re* 'broader') and superlative (*brad-ost* 'broadest').

The personal pronouns of OE were: *ic* 'I', *þu* 'thou, you, sg.', *heo* 'she', *he* 'he', *hit* 'it', *we* 'we', *ʒe* 'you, pl'., *hie/hy* 'they'. They agree with the main verb in person (first, second, third) and number (singular, plural). In addition to the familiar singular–plural number distinction, OE had two special 'dual 'pronouns: *wit* 'the two of us' and *ʒit* 'the two of you'. Like the nouns, the pronouns had case forms depending on their syntactic function. Personal pronouns and demonstratives also had three genders: feminine, masculine, and neuter. There were no gender distinctions in the plural.

The OE verbs had inflexions for the following grammatical categories:

(a) Tense: present and preterit, e.g. *fylle- fylde* '(I) fill-filled'; *riseþ-ras* '(she) rises-rose'.
(b) Person: first, second, third, e.g. *fylle-fyllest-fylleþ* '(I) fill-(you) fill-(he) fills'.
(c) Number: singular and plural, e.g. *fylle -fyllaþ* '(I)fill-(we) fill'.
(d) Mood: indicative, imperative, subjunctive, e.g. *fyllest-fyll(a)-fylle* '(you) fill-fill!-(should, would) fill'.

OE had two types of verbs. The 'weak' verbs used a dental suffix: *-de, -te, -ed, -od,* etc., to form the preterit and the past participle; *fylde* 'filled', *(Ʒe)-fylled* 'filled'. Strong verbs changed their root vowel, depending on tense and number: *rise, ras, rison* 'rise-(it) rose-(they) rose'. Modern English survivals of this group are called 'irregular' verbs. Verbs like *bind, choose, drink, eat, ride, run, see* belong here.

Unlike Modern English, which has a very elaborate set of complex verb forms, OE used such forms very sparingly. The verbs *willan* 'want' and *scullan* 'must' were used to express the future tense only occasionally, and the verbs *have* and *be* had not developed their auxiliary functions, as in *we have arrived, we had been traveling.* The predecessors of the Modern English modal verbs, e.g. *cunnan* 'know, be able', *magan* 'be able', *motan,* p.t. *moste* 'be allowed to', were more like the rest of the verbs in the language in that they had tense, person, number, and mood inflexions.

In terms of word-order, OE was, to a large extent, a 'verb-second' language: typically, the inflected verb in main clauses appeared in the second constituent position. The first position could be filled by the subject, by one or more adverbs, and even by an object, although noun objects normally followed the verb; statistically the order V–O was dominant. The subject could precede or follow the verb, depending on whether the first constituent position in the clause was filled and on whether the subject was a full noun or a pronoun.

> On þam twelftan dæge cristes acennednysse *comon* þa þry tungelwitegan…

> On the twelfth day of Christ's birth *came* the three astrologers…

With complex verb forms, the first verb occupied the second position, while the second part of the verb, an infinitive or a past participle, appeared at the end of the clause. Thus, the two parts of the verb formed a 'brace' enclosing all but the very first clause constituent.

> We *sceolon…* mid halgum mægnum þone eard *ofgan…*
> We *must…*with holy virtues this country *obtain…*

In subordinate clauses, the normal position of the inflected verb was at the end:

> Ac heo weron synfulle, þeah þe heo swa ne *wendon…*
> But they were sinful, though they so (did) not think.

The verb *do* was not used as an auxiliary verb in OE. Negative sentences were formed by attaching the particle *ne* 'not' to the left of the verb, and questions required inversion of the verb and the subject. Multiple negation was the norm in OE.

> *Ne* genealæceþ him *nænig* yldo ne sorh ne sar ne deaþ.
> Not approaches him no old age, nor sorrow, nor pain, nor death

Vocabulary and Word Formation

The recorded vocabulary of OE is estimated at approximately 30,000 words. Only about 3% of these were of non-Germanic origin. The number of Celtic borrowings such as *binn* 'bin', *torr* 'rock', is limited. The influence of Latin is most noticeable in religious and learned texts; *abbod* 'abbot', *cantere* 'cantor', *fers* 'verse', and *Læden* 'Latin' are some examples. The most significant outside addition to the OE vocabulary comes from Scandinavian. Although many Scandinavian loanwords were not recorded until after the Norman Conquest, OE is arguably the period during which about 1,000 words were borrowed from Scandinavian. This component of the vocabulary includes everyday words: Modern English words such as *bank, call, fellow, guess, law, leg, loan, score, skill, sky, skin, take* belong in this group. Scandinavian is the source of the pronouns *they, them, their,* and the function words *till* and *though.* There are also about

1,400 Scandinavian place names in the northern and eastern parts of England. These are place names ending in -*by* 'settlement' (*Carnaby, Rugby, Thirtleby*), -*thorpe* 'hamlet' (*Barleythorpe, Grimsthorpe, Fridaythorpe*), -*thwaite* 'clearing' (*Hampsthwaite, Hunderthwaite, Husthwaite*).

OE was extremely resourceful in enriching its vocabulary through derivation and compounding. New items could be derived by prefixation as in *a-sendan* 'send away, dispatch', *in-genga* 'invader', *un-wisdom* 'stupidity'. Frequently used OE suffixes are: -*ere*, as in *fisc-ere* 'fisher', -*estre*, as in *spinn-estre* 'female spinner', -*ing*, as in *cyn-ing* 'head of the kin, king'. Other highly productive OE suffixes were: -*dom, -had, -scipe, -nes(s)* for nouns, and -*ig, -isc, -sum, -en, -fæst, -lic, -leas, -ful* for adjectives. Compounding was another prominent vocabulary enrichment strategy. Some examples are: *heafod-mann* 'head-man, leader', *eorð-cræft* 'earth-craft, geography', *wid-sæ* 'wide-sea, ocean', *mild-heort* 'mild-hearted, merciful'. Metaphoric compounds, or *kennings*, are characteristic of the OE poetic tradition, e.g. *ban-cofa* 'chamber for bones, body', *gar-berend* 'sword carrier, warrior', *guðwudu* 'battle wooden piece, spear'.

References

Bosworth, Joseph. 1992. An Anglo-Saxon dictionary: based on the manuscript collections of Joseph Bosworth. Supplement by T. Northcote Toller, with rev. and enl. addenda, by Alistair Campbell. London, New York: Oxford University Press, c. 1921.

Campbell, Alistair. 1959. Old English grammar, Oxford: Clarendon Press.

Fulk, Robert D. 1992. A history of Old English meter. Philadelphia: University of Pennsylvania Press.

Healey, Antonette di Paolo (ed.) 1998. Dictionary of Old English Old English Corpus. Ann Arbor: University of Michigan Press. Electronic access: http://ets.umdl.umich.edu/o/oec/.

Hogg, Richard, M. 1992. A grammar of Old English. Oxford, UK; Cambridge, MA., USA: B. Blackwell.

——— (ed.) 1992. The Cambridge history of the English language: Vol. I, The beginnings to 1066. Cambridge: Cambridge University Press.

Lass, Roger. 1994. Old English: a historical linguistic companion. Cambridge: Cambridge University Press.

Mitchell, Bruce. 1985. Old English syntax, 2 vols. Oxford: Clarendon.

Mitchell, Bruce, and Fred C. Robinson. 1992. A guide to Old English, 5th edition. Oxford, UK; Cambridge, MA., USA: Blackwell.

Nielsen, Hans Frede. 1998. The continental backgrounds of English and its insular development until 1154. Odense: Odense University Press.

Quirk, Randolph, and C.L. Wrenn. 1994. An Old English grammar, with a supplemental bibliography by Susan E. Deskis. DeKalb: Northern Illinois University Press.

Roberts, Jane Annette, and Christian Kay with Lynne Grundy. 1995. A thesaurus of Old English, 2 vols. London: King's College London Medieval Studies 11.

DONKA MINKOVA

See also **Great Britain; Indo-European 2: Germanic Languages**

Old French

'Old French' refers to the set of dialects spoken in northern and central France from the ninth to the fourteenth centuries. The term is most commonly identified, especially in the later stages, with *francien*, the speech of the Île-de-France region surrounding Paris, because of that variety's central geographic position and its association with the royal court and the burgeoning political, economic, and cultural importance of the capital.

The birth of Old French (henceforth OF) is traditionally linked with the *Serments de Strasbourg* (AD 842), a document containing oaths sworn by the troops of two of Charlemagne's grandsons concerning the partition of his empire. This document contains structures (particularly a new future form) sufficiently different from Classical Latin (henceforth CL) to allow the conclusion that a new language has emerged (although any firm date is clearly artificial—the changes had to have been in effect for some time prior to their appearance in the *Serments*). The early OF period extends from the ninth to the early twelfth century, and constitutes a period of rapid linguistic change and considerable regional diversity. The later OF period (mid-twelfth to fourteenth century) sees the establishment of the University of Paris and the law courts, the flourishing of literary works (by no means limited to the dialect of the Ile-de-France), and the increasing prestige and expansion of *francien*.

How might one characterize linguistically this new language? (We must bear in mind that any summary of five centuries of evolution is inevitably far from complete—we concentrate here on one representation of

the 'classic' forms of the twelfth century and on the major differences leading from Latin through to Modern French [MF].) In its pronunciation, OF lost the distinction between long and short vowels characteristic of Latin; often modified stressed vowels (BREVIS /brĕwis/ > OF *brief* /brjɛf/ > MF *bref*[1] 'brief', HORA /hoRa/ > OF *eure* /ewrə/ > MF *heure* 'hour'); deleted many unstressed or final vowels (except /a/ and those vowels following a final group of consonants [final /a/ becomes /ə/]: MALUM > *mal* 'bad', ARBOREM > *arbre* 'tree', GUTTA > *goute* 'drop'); introduced the affricates [tʃ dʒ ts dz] through palatalization, as in OF *ch*a*lt* 'hot', *jorn* 'day', *cent* 'hundred', *onze* 'eleven', respectively; began a long-term process of vowel nasalization culminating in forms such as MF *un* [œ̃] 'one' < UNUM, *chien* [ʃjɛ̃] 'dog' < CANEM, *pont* [põ] 'bridge' < PONTEM; and deleted or weakened consonants between vowels, as in VITA > *vie* 'life', SAPONEM > *savon* 'soap'.

In word structure, the changes were no less dramatic. The loss of final vowels had profound effects on many grammatically relevant suffixes. Among the nouns and adjectives, the original CL system used suffixes to distinguish five separate classes of nouns, three genders (masculine, feminine, and neuter), two numbers (singular and plural), and six cases (nominative, genitive, dative, accusative, vocative, ablative, each with various syntactic roles indicating, among other things, subject [nominative case], direct object [accusative] and indirect object [dative], possession [genitive], terms of address [vocative], and so on). These were reduced in OF to three classes of nouns, two genders, two numbers, and two cases. With the disappearance of the final vowels, the suffixes disappeared in large part, so that the syntactic distinctions originally indicated by these suffixes were also directly affected. As a result, we see compensating developments in the increased use of prepositions (e.g. *le livre de Cicéron* 'the book **of** Cicero = Cicero's book' rather than LIBER CICERONIS), and OF word order patterns become more constrained than in CL (where, because suffixes indicated the words' function, words could appear in virtually any order depending on stylistic factors [CANEM ['dog–accusative'] HOMO [man—nominative] VIDET > *l'homme voit le chien* 'The man sees the dog']). The determiner system also underwent radical change, with great increases in the use of demonstratives (e.g. OF *cest* 'this', *cel* 'that') and particularly articles (e.g. OF *li, le, la, les* 'the (sg. and pl.)', *uns, un, une* 'a (masc. and fem.)').

Declensional classes for nouns are illustrated below for *murs* 'wall', *rose* 'rose', *ber* 'baron', and *none* 'nun', respectively.

OF declensional classes (major types only, including definite articles; imparisyllabic forms have a different number of syllables in the nominative singular as compared to the remaining forms; 'oblique' refers to an amalgam of all nonnominative cases):

	Masculine	Feminine	Double (Imparisyllabic)	Stems
Nominative singular	li murs	la rose	li ber	la none
Nominative plural	le mur	la rose	le baron	la nonain
Oblique singular	li mur	les roses	li baron	les nonains
Oblique plural	les murs	les roses	les barons	les nonains

(Thus, *La none li baron veit.* 'The nun sees the baron.')

In the verb system, equally striking losses occurred. Many distinctions (passive forms, for example) disappeared completely. A new future tense arose, built with the infinitive and suffixes based on HABERE 'to have': CANTABO (the older form) > CANTARE HABEO 'I have to/will sing' > OF *chanterai* 'I will sing.' Verb endings also eroded, so that person/number distinctions (I, you, he/she/it [sg.]; we, you, they [pl.]), formerly indicated by suffixes, were also threatened. As a result, pronoun use gradually increased, especially in the singular as a way of preserving the former differences.

The form of OF words was significantly more variable than in the modern language. The effects of earlier sound change often resulted in different forms of noun and verb stems that were regularized subsequent to the OF period (OF *truef—trovons* > MF *trouve—trouvons* 'find' [1 sg.—1 pl.]; *serf—sers* > MF *serfs—serfs* 'serf' [sg.—pl.]; OF *larc—large* [masc.—fem.] > MF *large* 'wide'). Fusion of unstressed forms is also widespread: *a le > au* 'to + the (sg.)', *en les > es* 'in + the (pl.)', *je le > jel* 'I + him/it', *ne les > nes* 'neg + them', *qui me > quim* 'who + me', *si me > sim* 'if + me', and numerous others.

In its sentence structure, OF still permitted relatively free word order compared to the modern language, but certain patterns, particularly those where the verb is in the second position in the sentence, began to predominate: subject–verb–object: *Li vilains apele son fil.* 'The peasant calls his son.' [*Fabliaux* 3, l. 39]; the frequent object–verb–subject (– X): *Ses barons fist li rois venir.* 'The king had his barons come.' [*Le roman de Renart*, l. 1807]; *De venoison ont grant plente.* 'They

---

[1] The symbol '>' indicates the direction of development: Latin BREVIS becomes (>) OF *brief*. Conversely, '<' indicates the source: OF *chien* comes from (<) Latin CANEM.

have a lot of venison.' [*Tristran*, l. 1773]; and subject–object–verb: *Li rois Tristran menace.* 'The king threatens Tristran.' [*Tristran*, l. 770]. Given the disappearance of the CL passive verb forms, it is not surprising that replacements of these CL constructions appeared based on *estre* 'to be' plus the past participle (*La pucele fut donc pendue.* 'They hanged the young girl' [*Vie de sainte Marguerite*, l. 1]); on the increased use of pronouns such as *on/l'en* (*On me desrobe en vostre terre.* 'I am being robbed in your land.' [*Fabliaux* 11, l. 191–2]); or on the expansion of constructions with the reflexive pronoun *se* (*Carles se dort.* 'Charles falls asleep.' [*Chanson de Roland*, l. 724]). OF sentence patterns, in other words, remind us much more of the modern language than of CL.

In vocabuary, the Latin origins of OF are clear —the great majority of lexical items descend directly from Latin, although often from informal or popular rather than classical speech. In addition to popular items, which have become the norm (TESTA 'jug' > OF *teste* 'head'; CABALLUS 'nag, packhorse' > OF *cheval* 'horse'; GENUS 'knee' — GENUCULUM [vernacular diminutive form] > OF *genoil* [Modern French *genou*] 'knee'), we see innovative use of many suffixes (e.g. OF *-age* used to indicate taxes of various types: *abeillage, arivage, cheminage, melage* [on bees, docking, roads and apples, respectively]; Germanic loans from the early Frankish conquerors (*biere* 'coffin', *helme* 'helmet', *honte* 'shame'); learned words entering from religious or legal texts (*avaricieux* 'avaricious', *crestiien* 'Christian', *testimonie* 'testimo-

ny'); and importations from a variety of other sources (Celtic *chemin* 'road', *if* 'yew tree'; Arabic *alchimie* 'alchemy'; Greek *eglise* 'church'; Occitan *amour* 'love' [notably via the influence of the Troubadour poets], and many others). Nonetheless, the unmistakably major lexical source is Latin.

But despite this parentage, OF is clearly French, no longer a 'corrupted' Latin, and has distanced itself from its origins much more than such well-known Romance relatives as Italian, Occitan, Catalan, or Spanish. While not easily comprehensible to modern readers, it leaves a stong impression of familiarity, and should inspire us as a source of great cultural and linguistic richness.

References

Buridant, Claude. 2000. Grammaire nouvelle de l'ancien français. Paris: SEDES.

Einhorn, E. 1974. Old French. A concise handbook. Cambridge, New York: Cambridge University Press.

Harris, Martin. 1978. The evolution of French syntax. A comparative approach. London, New York: Longman.

Kibler, William. 1984. An introduction to Old French. New York: The Modern Language Association of America.

Pope, Mildred. 1934. From Latin to Old French with especial consideration of Anglo-Norman. Manchester: Manchester University Press.

Price, Glanville. 1971. The French language: present and past. London: Edward Arnold.

DOUGLAS C. WALKER

See also **France; Indo-European 4: Romance; Latin**

Old High German

The German language is estimated to be about 1,500 years old based on the earliest written sources, though as a spoken language it is almost surely older. Old High German is the traditional designation of the German language in its earliest stage, roughly between the years 750 and 1050 CE (some scholars use the period 600–1100).

German is a member of the Germanic subfamily of the 'centum' branch of the Indo-European language family. The linguistic features that set Germanic off against other Indo-European languages are several, principal among which are the First Sound Shift (also called the Germanic Sound Shift, the First or Germanic Consonant Shift, or Grimm's Law), Verner's Law, the

fixing of main intonational stress on the initial syllable of the word, and a variety of vowel changes.

The origins of the Germanic peoples—the speakers of the ancestor language Proto-Germanic—are cloaked in mystery, as indeed are the origins of the Indo-Europeans. We are relatively certain that sometime prior to 1000 BCE Germanic-speaking people occupied their primeval home (*Urheimat*) in what is today the area comprising southern Sweden and Norway, Denmark, and northern Germany. Between 1000 BCE and 500 BCE, at least some of the Germanic tribes began to move away from their original home, migrating farther north or striking out to the south and the east. The reason for this migration was doubtless due in

part to inundation of the *Urheimat*, much of which is covered today by the relatively shallow North Sea.

Out of this *Völkerwanderung* ('migration of peoples') arose the traditional classification of the Germanic languages into East, North, and West Germanic. German belongs to the latter group, as do English, Dutch, Frisian, and Low German. We know nothing for certain about the people who occupied this territory before the Germanic tribes arrived. We know nothing about the early contacts between the Germanic intruders and the autochthonous inhabitants of the *Urheimat*, nor do we know what language(s) the latter spoke. The general rule of thumb in historical linguistics is that the greater the degree of contact between languages, the greater the amount of language change. The changes in pronunciation and vocabulary that Germanic languages underwent (the First Sound Shift and others) were extensive compared with more conservative Indo-European languages. Moreover, approximately one-third of the vocabulary of the Germanic languages is from other than Indo-European stock. Therefore, we assume that contacts with the autochthonous inhabitants were deep and extensive—and that the languages of these speakers had many fricative consonants (the source of English *f, th, s, sh*).

While there are competing representations for illustrating the relationships among the Germanic family depending on which criteria are given priority, the most widely used and traditional genetic classification of the Germanic languages is a tripartite one into East, North, and West. The family tree (stammbaum) of Proto-Germanic can be represented (and simplified) as follows. (The terms 'Ingvaeonic' and 'Istvaeonic' refer to the names of Germanic tribes; 'North Sea Germanic' and 'Weser-Rhine Germanic' are used almost interchangeably with 'Ingvaeonic' and 'Istvaeonic'. Old Low Franconian gave us Dutch, and Old Saxon was the ancestral language of modern Low German, which is spoken today in northern Germany.) (See Figure 1.)

The principal linguistic change that set Old High German apart from its other West Germanic sibling languages was the Old High German Consonant Shift, also called the Second Consonant Shift or the Second Sound Shift. This, along with the First Sound Shift differentiating Germanic from Indo-European, is one of the most illustrious sound changes in Germanic linguistics. It affected the voiceless stops /p t k/, changing them according to the phonetic environment into affricates /pf ts kx/ or fricatives /f s x/, the latter of which could occur as both long (geminated) /ff ss xx/
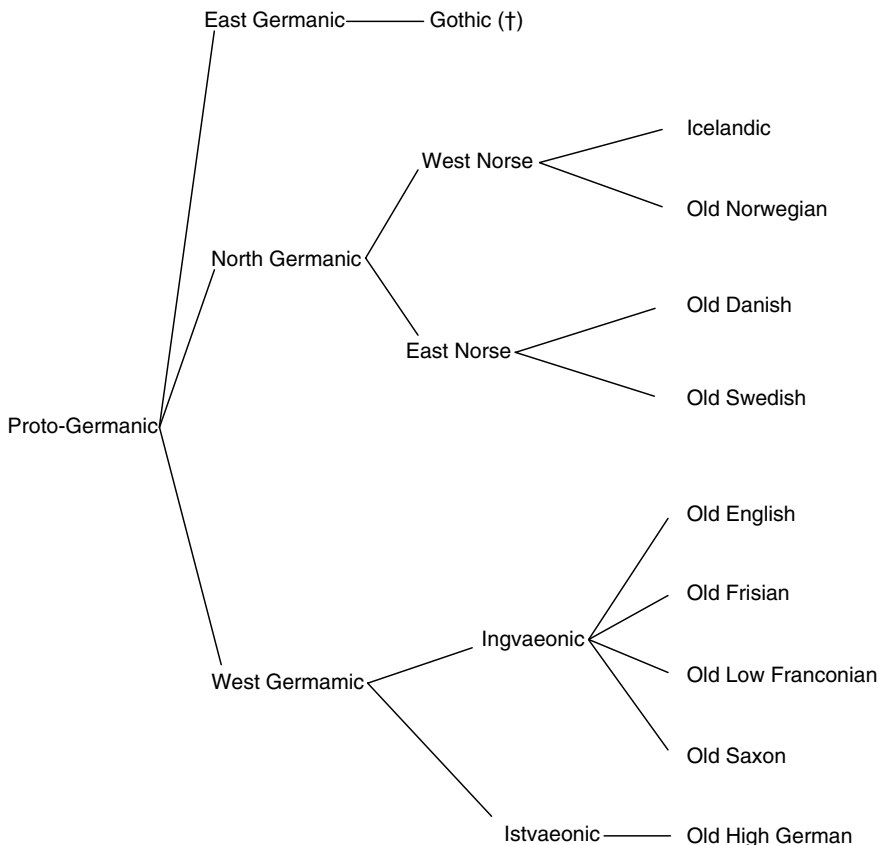
Figure 1

and short /f s x/. In general outlines, the Old High German Consonant Shift went as follows:

Pre-Old High German		Old High German	Phonetic Environment
p t k	>	pf ts kx	Word-initially and following m, n, l, r
	>	ff ss xx	Medially following a short vowel
	>	f s x	Medially following a long vowel and word-finally

We can see the effects of the Old High German Consonant Shift from a comparison of modern German words with their English cognates:

Modern German	English
Pfund	pound
Pfeffer	pepper
Wasser	water
machen	make
Zeit	tide
helfen	help
auf	up
kochen	cook

An additional sound change of Pre-Old High German *d* to Old High German *t* is considered by some linguists to be part of the Old High German Consonant Shift. Its effects are seen in cognates such as English *do, day, ride* vs. German *tun, Tag, reiten.*

The designation 'High' has two meanings in the context of High vs. Low German. On the one hand, it is a purely linguistic designation for those dialects of West Germanic that underwent any part of the Old High German Consonant Shift. Its original meaning in this context, however, referred to altitude: those dialects of German that did not undergo the Old High German Consonant Shift are located in northern Germany on the flatlands bordering the North and Baltic Seas—'low lying'in other words, hence 'Low' German—and the dialects of German that did undergo the Old High German Consonant Shift are located in the higher regions of southern and central Germany, Switzerland, and Austria, thus 'High German'. One occasionally encounters the use of 'High' to mean 'good, correct, elevated, superior, grammatical' and 'Low' to mean the opposite, linguistically, of those qualities. One does not encourage this usage.

Old High German is divided into Central German and Upper German dialects. Upper German consists of those dialects spoken in southern Germany, Switzerland, and Austria. The principal Upper German dialects of Old High German are Alemannic, spoken in present-day Switzerland, and Bavarian, spoken today in the German state of Bavaria and in Austria. Central German dialects are spoken in an east–west band lying between Low German in the north and Upper German in the south. The principal Central German dialects are Franconian (Middle Franconian, Rhine Franconian, East Franconian) and East Central German, which overlaps with Thuringian, Upper Saxon, and Silesian.

Upper German dialects are characterized by greater completeness in the effects of the Old High German Consonant Shift. It is assumed therefore that the Old High German Consonant Shift began in the far south and spread toward the north: the underlying idea is that sound change goes further toward completion in the area it has been around in longest. Most Upper German dialects have all of the changes stated above as comprising the Old High German consonant shift. Central German dialects show more mixed results. Central German dialects are most consistent in the shift of *k* to *x* and less consistent in the shifts of *t* to *s* and *p* to *f.* Thus, where Upper German dialects will consistently have fricatives in **ich** 'I', *dorf* 'village', and *das* 'that', Central German dialects can be found with **ich**, *dorf, dat* and **ich**, *dorp, dat.*

The distinctive sounds of Old High German are:

Vowels

Short		Long		Diphthongs	
i	u	ī	ū	ei	ou
e	o	ē	ō	ie	uo
a		ā			

Consonants

		Labial	Dental	Palatal	Velar	Glottal
Stops	vl.	p	t		k	
	vd.	b	d		g	
Fricatives	vl.	f	s̲	ṣ	x	
Affricates	vl.	p^f	t^s		k^x	
Nasals		m	n			
Liquids			l		r	
Semivowels		w		y		h

([s̲] and [ṣ] are respectively dorsal—formed with the blade of the tongue against the alveolar ridge—and apical—formed with the tip of the tongue against the alveolar ridge. They have different historical sources: [s̲] is created by the shift of *t* to *s* in the Old High German Consonant Shift; [ṣ] continues Indo-European *s*).

The back vowels [u o a ū ō ā] had front variants [ü ö æ ǖ ȫ ǣ] before syllables containing a high front vowel or semivowel [i ī y], thus *gesti* 'guests', *skôni* 'beautiful', and *ubil* 'evil' were phonetically [gesti skoni übil]. These are the so-called 'umlaut' sounds of German (*Gäste, schön, übel*).

The German language is traditionally divided into four periods: Old High German (750–1050); Middle High German (1050–1350); Early New High German

(1350–1650); and New High German (1650–present). The division between Old High German and Middle High German is dated from the reduction of the unstressed vowels [i e a o u] into a single sound schwa [ə]. Since this did not occur in a single day, the transition from Old High German to Middle High German is variously placed between 1050 and 1100, at which time the reduction of unstressed vowels to schwa was complete except in remote and isolated dialects. From the beginning of the Common Era, Old High German was in contact with Romans during the expansionist phase of the Roman Empire. Lexical borrowings were heavy, and their semantic domains reveal much about the nature of these early contacts: 'wine' (Old High German *wîn*, Latin *vinum*), 'arrow' (Old High German *pfîl*, Latin *pilum*).

The earliest written records of the Old High German language are words and fragments written in runes. Somewhat later, after the Germanic tribes had been Christianized, we find Old High German glosses interspersed in Latin texts. In the *Vocabularius Scti Galli*, slightly later than 765 CE, we find *sapiens wizzo* (in normalized Old High German spelling), *scitus wiser, fortis stark*. Lengthier texts are found beginning in the ninth century, many of them translations into Old High German of religious writings originally in Latin: Isidor's diatribe *Contra Judaeos*, the *Weissenburg Catechism*, the gospel harmonies *Tatian* and *Otfrid*, and the translation of Boethius made by Notker, an especially talented monk from the monastery of Saint Gall in Switzerland.

Typical of the genre of most early Old High German writing is the Lord's Prayer from St. Gall:

> Fater unsêr, dû bist in himile, wîhi namun dînan, kweme rîhhi dîn, werde willo dîn, sô in himile sôsa in erdu. brôd unsêr emezzihig gib uns hiutu, oblâz uns skuldi unsêro, sô wir oblâzêm uns skuldîgêm, enti ni unsih firleiti in khorunka, ûzzer lôsi unsih fona ubile.

Typical too were 'How to say it in Old High German' booklets such as the 'Conversations from Paris', written presumably for monks literate in Latin for traveling in Germany (the unnormalized spelling of Old High German points clearly and amusingly to the French background of the monk who wrote the thing):

> Gueliche lande cumen ger? (de qua patria?) 'What land do you come from?'
> Guer is tin erro? (ubi est senior tuus?) 'Where is your master?'
> Ne guez. (nescio.) 'I don't know.'
> Gimer min ros. (da mihi meum equum.) 'Give me my horse.'

Of particular cultural and historical interest are the 'Oaths of Strassburg', which date from 843. These are oaths of allegiance intended to end a fratricidal rivalry between the kings Louis and Charles of the Franks, sons of Charlemagne. Because the eastern Franks spoke Old Low German and the western Franks Old French, the oaths were composed in both languages, thus giving us linguistic information as useful for a student of the history of the French language as for a student of the history of German.

There is in Old High German a certain amount of original composition not derived from Latin translations and more substantial than 'give me my horse'. The *Muspilli* is a religious poem of some length that is strikingly original, what we have of it, and in the later Old High German period there are *Memento Mori* and *The song of Ezzo*.

It must be admitted that Old High German literature is inferior in comparison with the contemporaneous literature of Old English (*Beowulf*) or even its close relative, Old Saxon (*Heliand*). It is vastly inferior to the rich medieval literature of Middle High German with its courtly epics, its *Nibelungenlied*, its poetry. Old High German literature is not equal to the slightly later literature of Old Icelandic. There is very little original in Old High German literature, much that is derivative.

However, its literary deficiencies should not detract from its linguistic importance. It was the language of the Second Sound Shift, the Old High German Consonant Shift, which along with the First Sound Shift (Grimm's Law) and Verner's Law is one of the great defining phonological events in the Germanic family of languages. The Old High German Consonant Shift—its causality, its spread, how it happened—has provided and continues to provide historical linguistics with grist for its theoretical mill. The importance of a language must never be judged by its literature alone.

References

Barber, C. Clyde. 1951. An Old High German reader. Oxford, Blackwell.

Braune, Wilhelm. 1987. Althochdeutsche Grammatik. Tübingen: Max Niemeyer (numerous editions).

Hawkins, John A. 1987. 'German' The world's major languages, ed. by Bernard Comrie. New York: Oxford University Press.

Lockwood, W.R. 1968. Historical German syntax. Oxford: Clarendon Press.

Murdoch, Brian O. 1983. Old High German literature. Boston: Twayne Publishers.

Priebsch, R., and W.E. Collinson. 1962. The German language. London: Faber and Faber.

Waterman, John T. 1976. A history of the German language. Seattle: University of Washington Press.

Wells, C.J. 1985. German: a linguistic history to 1945. Oxford: Clarendon Press.

Wright, Joseph. 1906. An Old High German primer. Oxford: Clarendon Press.

ROBERT D. KING

See also **Germany; Grimm, Jacob; Indo-European 2: Germanic Languages**

Old Irish

Old Irish is a now extinct Indo-European language belonging to the Insular Celtic branch of the Celtic language family. Insular Celtic is that branch of Celtic that is (relatively speaking) indigenous to the British Isles. Old Irish is the direct ancestor of present-day Irish, as well as of Scots Gaelic and Manx.

Within earlier Insular Celtic, it is possible to distinguish two major dialects, known respectively as Q-Celtic (sometimes called Goidelic or Gaelic) and P-Celtic (sometimes called Brythonic or Britannic). These labels have been retained by linguists to classify the later distinct Celtic languages. Q-Celtic is so called because it shows /k/ as the descendent of an original Indo-European /*kʷ/. (The name 'Q-', rather than 'K-', Celtic originates in Latin spelling.) In contrast, P-Celtic shows /p/ as the descendent of /*kʷ/. Irish, along with Scots Gaelic and Manx, belongs to the Q-Celtic group, while the P-Celtic group contains Welsh and Cornish, as well as Breton. We may contrast the development of the inherited word for 'four', which in Old Irish became _cethair_ but in Old Welsh became _petguar_.

The earliest surviving evidence for Q-Celtic comes from inscriptions written in the native Ogham alphabet, which date from the fourth to seventh centuries CE. These consist primarily of proper names, and so the amount of grammatical information that they can provide is limited; however, they do provide information about the sound system and word structure of Q-Celtic in the immediately pre-Old-Irish period.

The Old Irish period proper is normally considered to fall between 600 and 900 CE, and it is from this period that the earliest literary evidence for Irish comes. Within the Old Irish period are sometimes further distinguished an Archaic Old Irish period (prior to the 700s CE) and a Later (or Classical) Old Irish period (the 700s and 800s CE). Ogham writing, incidentally, continued to be taught until around the seventeenth century, but its role after the inscriptional period of Q-Celtic was greatly restricted and most Old Irish material is thus written using a form of the Roman alphabet.

The corpus of contemporary texts that represents Old Irish is relatively limited in both size and genre. The most important texts, dating from the Classical Old Irish period, consist of marginal notes (or 'glosses') on Latin manuscripts. Of these, the most important are the Würzburg glosses, dated mainly to around 750 CE, the Milan glosses, normally considered to date from sometime before 825 CE, and the St. Gall glosses, dating in manuscript from around 845 CE but containing copies of earlier material. To these texts can be added a number of further sources consisting of additional glosses as well as a small number of continuous texts such as poems. In addition to the Classical Old Irish texts, there also exist a small number of older texts from the 600s CE.

It is generally agreed that further Old Irish material is also preserved in later manuscripts, and, indeed, it has been argued that some of this material illustrates a period of Old Irish that pre-dates the earliest contemporary manuscripts. Certainly, some of these texts do show differences in comparison with other texts of the same date, notably in word order. However, this is a disputed area, and it has been argued convincingly that at least some of this variation may be due more to genre than to date. The texts may also have come under the influence of Latin models, and some of their distinctive characteristics may in part reflect interference from Latin. Thus, the contemporary Old Irish manuscripts still form the main, and safest, basis for the linguistic description of Old Irish, although scholars vary as to how far they will admit evidence from the later manuscripts.

The Old Irish texts that have come down to us are linguistically very homogeneous and show rather little evidence of dialect variation. The classical view has been that what variation does exist is due to differences in date, although it has been tentatively suggested that these differences do, in fact, reflect dialectal variation. Certainly, the relative homogeneity of the Old Irish sources should not be taken to indicate that some degree of dialect variation did not exist within Old Irish; indeed, it would be unusual if this were the case, since the language was spoken in the whole of Ireland and in a large part of Scotland. We do not know how small or substantial these dialect differences may have been in the spoken language. As regards the homogeneity of the literary sources, it should be remembered that they were written by scholars in monastic communities, and some degree of dialect mixing, and possibly informal standardization, will undoubtedly have taken place.

The vocabulary of Old Irish is predominantly of Q-Celtic origin, but also contains substantial loan material, primarily from P-Celtic and from Latin. Some of the Latin loans have undergone sound changes that suggest they were borrowed indirectly via P-Celtic, while others seem to have been borrowed directly.

In structural terms, Old Irish has a number of features that are interesting from a historical and comparative perspective.

On the one hand, the evolution of Old Irish to modern Irish shows trends toward the breakdown of complex systems of functional endings (inflexions), simplification, and the atomization of meaning-carrying elements that are paralleled in other language groups—for instance, in the development from Latin to modern Romance languages such as Italian and French.

In the noun, for instance, Old Irish maintained a complex set of inflexions: for example, it had special dual forms for pairs of things, as well as a set of five case endings indicating grammatical relations such as subject, object, and possession. The same was also true of the adjective, which agreed with the noun that it modified. Although already largely obsolete in Old Irish, owing to the emergence of a fairly fixed word order, this system of endings was to be retained well into the modern Irish period (up to the 1600s CE). However, it has since broken down substantially, and must already have been breaking down at an earlier date in the spoken language. Some modern Irish dialects have practically abandoned noun inflexion, and even the more conservative standard written language uses a reduced set of endings.

The Old Irish verb, too, was exceptionally complex, and has consequently received much attention from linguists. The basic distinction was between simple verbs and complex verbs, that is, those that were formed from simple verbs by the addition of various preverbal elements. For instance, from *beirid* ('bear') could be derived compound verbs such as *do·beir* ('give'), *for·beir* ('grow'), *as·beir* ('say'), and *tremi·beir* ('transfer'). In addition to these compounding preverbal elements, other preverbal affixes existed: for example, *ro* was used to distinguish, among other things, between the perfect and simple past (contrast English 'I have run' and 'I ran'). Each of the two classes of verb had, in turn, two full sets of personal endings, which were used in parallel to one another, depending on the grammatical context: in the simple verb, they were known as absolute and conjunct endings, and in the compound verb as prototonic and deuterotonic endings. In addition, there were special endings in the simple verb to indicate a verb's participation in a relative clause. Furthermore, as with other languages such as Latin, pronoun subjects were not expressed, and the meaning was carried by endings (contrast the unchanging *bear* in English 'I bear' and 'you bear' with Old Irish *biru* ['I bear'] and *biri* ['you bear']). One of the main changes distinguishing Old Irish from later periods of Irish is the breakdown, and consequent simplification, of this complex verbal system, notably the disappearance of the two sets of absolute and conjunct endings and the near disappearance of the prototonic and deuterotonic distinction.

On the other hand, Old Irish also has features that are characteristic of itself and of the wider Celtic language family, in contrast with other Indo-European languages.

In common with the other Celtic languages, Old Irish has Verb–Subject–Object as its basic word order. The Celtic languages are the only Indo-European languages to have this as their basic word order. There has been some discussion about the emergence of this word order in Old Irish. One suggestion has been that it is the consequence of generalizing the order that resulted from suffixed pronouns (see below).

In Old Irish, as in other Celtic languages, the boundaries between the different language levels of sentence structure, word structure, and sound structure are not discrete. For instance, the initial sounds of Old Irish words can change in certain contexts (so-called 'mutations'). These changes were originally triggered by the final sound in the preceding word, but a number of these sounds later disappeared and the pattern of mutation was also extended by analogy to other contexts. Some mutations thus came to be used to mark grammatical relationships, rather than being the result of local phonetic influence. This is the case, for example, in certain kinds of relative clause, where the main verb of the relative clause undergoes mutation. Thus, sentence function triggers a change in wordform, which was originally phonetically motivated.

A further peculiarity of Old Irish is the frequency with which certain grammatically interrelated words within a sentence or clause are fused into a single word unit, where, in contrast, other languages (such as English) would normally use two or more free-standing words. One example of this is the situation where the object of a verb is a pronoun. In Old Irish, pronoun objects were often suffixed to the verb, rather than standing on their own as separate words–for example, *beirthi̱* ('bears it') or *iurru̱s* ('will wound them'). Following one of the preverbal elements referred to earlier, a pronoun could actually become infixed within the verb—for example, *ro·m̱·gab* ('has taken me'). In contrast to this Old Irish tendency, pronouns in modern Irish may also be free-standing, and the use of this option is becoming increasingly frequent, at the expense of the suffixed pronoun construction. A related phenomenon, which modern Irish has retained, is the case of the so-called 'conjugated prepositions', in which forms of the pronoun are also suffixed to prepositions, resulting in 'one-word' expressions such as *lat* ('with you', from *la* = with and *-t* = you).

Aside from its importance as a crucial chapter in the history of modern Irish and the other Q-Celtic languages, Old Irish is linguistically important more

generally for two main reasons. Firstly, from a historical perspective, it provides further evidence for the universality of certain principles of language change. Secondly, it is interesting typologically, in that it exhibits a number of features not normally found in Indo-European languages outside the Celtic language family.

References

Ahlqvist, Anders. 1988. Remarks on the question of dialects in Old Irish. Historical dialectology, regional and social, ed. by Jacek Fisiak. Berlin, New York: Mouton de Gruyter.

Lewis, Henry, and Holger Pedersen. 1937. A concise comparative Celtic grammar. Göttingen: Vandenhoeck and Ruprecht.; 3rd edition with supplement, 1989.

Mac Eoin, Gearóid. 1993 Irish. The Celtic languages, ed. by Martin J. Ball and James Fife. London, New York: Routledge.

McCone, Kim. 1996. Prehistoric Old and Middle Irish. Progress in Medieval Irish studies, ed. by Kim McCone and Katharine Simms. Maynooth: Department of Old Irish, St. Patrick's College.

———. 1997. The Early Irish verb. Maynooth: An Sagart.

Sims-Williams, Patrick. 1984. The double system of verbal inflexion in Old Irish. Transactions of the Philological Society 82.

Strachan, John. 1904. Old Irish paradigms and selections from the Old Irish glosses. Dublin: School of Irish Learning ; 4th edition, revised by O. Bergin, Dublin: Royal Irish Academy, 1949.

Thurneysen, Rudolf. 1946. A grammar of Old Irish. Dublin: Dublin Institute for Advanced Studies.

ANDREW WILSON

See also **Celtic Languages**

Old Japanese

Old Japanese is the oldest attested representative of the Japonic (Japanese-Ryukyuan) language family. There are two major varieties of Old Japanese: (1) Western Old Japanese (seventh to eighth century CE), a language based on the dialect of Asuka and Nara regions (roughly corresponding to modern Nara prefecture); and (2) Eastern Old Japanese (eighth century CE), a dialectal continuum located roughly in the southern part of the modern Chûbu and Kantô regions. Western Old Japanese was the basis of the literary language of the time, and consequently there are many more extant Western Old Japanese texts than Eastern Old Japanese texts. Furthermore, it is quite apparent that the westernmost dialects in the Eastern Old Japanese dialect continuum were considerably influenced by the Western Old Japanese standard, and this influence gradually diminishes from west to east. There is a widespread but mistaken opinion that Western Old Japanese is practically identical to Middle (Classical) Japanese. On the contrary, both languages are very different, with a number of important distinctions found in pronunciation, vocabulary, and grammar. Furthermore, Western Old Japanese is not even a direct predecessor of Middle Japanese: both are based on geographically close but not identical dialects.

Sources: All Eastern Old Japanese and most Western Old Japanese texts are poetry, although Western Old Japanese also has two texts that are written in prose. The major noninscriptional texts are: poetry from the 'Kojiki (Records of Ancient Matters)'

(712 CE), the 'Nihonshoki (Annals of Japan)' (720 CE), and the 'Fudoki (Gazeteers)' (713–737 CE); poetic anthology 'Man'yôshû (Collection of myriad leaves [of words])' (c. 759 CE), prose texts 'Senmyô (Imperial Edicts)' (seventh to eighth century), and 'Norito (Liturgies)' (seventh to ninth century). There are several important inscriptions as well, the largest of them being a poetic one: 'Bussoku seki uta (Songs of the Buddha's footprint)' (c. 755 CE).

Writing system: Western Old Japanese uses the writing system known as 'man'yôgana' (the 'Man'yôshû' writing). Man'yôgana is a syllabic system of writing, where Chinese characters are used phonetically (to represent syllables), although they can also be used logographically (to represent words or concepts). When Chinese characters are used both phonetically and logographically, there is a clear tendency to write roots or stems logographically and suffixes and particles phonetically; however, due to the syllabic nature of writing, the exact boundaries between stems (especially verbal stems) and suffixes may not be reflected in writing. Man'yôgana can be subdivided into two major types according to the historic stage of the underlying Chinese pronunciation of the characters on which man'yôgana is based. The later type is based on the Late Middle Chinese character readings of the eighth century, and it is used exclusively in the Nihonshoki. The earlier type is based on the Early Middle Chinese character readings of the sixth to seventh century; most likely it was not coined

directly in Japan on the basis of these readings, but was borrowed from Korea. All other Western Old Japanese texts use this earlier type of the man'yôgana. Eastern Old Japanese also uses this type of Western Old Japanese writing system, and it is necessary to note that all Eastern Old Japanese texts are essentially written in Western Old Japanese orthography.

Phonology: Western Old Japanese has the following consonants: *p, b [ᵐb], t, d [ⁿd], k, g [ᵑg], m, n, s, z [ⁿz], w, y, r*. No consonant is possible in the syllable-final position; thus, only syllables of C(onsonant)V(owel) and V structure are found in the language. Consonants *b, d, g, z,* and *r* occur only in word-medial position in native vocabulary. There are no sequences *wu and *yi in Western Old Japanese, but the rare exceptions to the general syllable template of [C]VCVCV…, like *kai* 'oar', suggest that at least the sequence *yi existed once at the pre-Western Old Japanese stage. The Eastern Old Japanese consonant system seems to be identical to Western Old Japanese.

Western Old Japanese vowels include the following vowels (Yale notation is given with the most likely phonetic values in brackets): *a [a], yi [i], u [u], iy [ï], ye [e], o [ə], wo [o]*, and one diphthong *ey [əy],* which was previously considered to be a unit vowel *[ɛ]*. The contrast between vowels *yi/iy, ye/ey*, and *o/wo* that later merged as /i/, /e/, and /o/ is not found in every possible position even in Western Old Japanese. Thus, there is no contrast in initial position for any of the three pairs; there is no contrast between *yi/iy* and *ye/ey* after coronal consonants (such as *t, d, s, z*), and a suggested contrast between *po* and *pwo* is questioned by some linguists, although it is likely that there is some supportive evidence for it in the earliest texts. In all cases where there is no contrast, it is customary to write just *i, e, o*.

The set of Eastern Old Japanese vowels is markedly different: it includes only five vowels: *a, i, u, e, o*, although there is the possibility that one or two additional vowels can be hidden behind the Western Old Japanese spelling system. In any case, vocalic correspondences between Western Old Japanese and Eastern Old Japanese are not straightforward, indicating that a parent language of both used to have a system somewhat different from its descendants.

Data on Western Old Japanese accent system are controversial: it has been suggested that in the 'Nihonshoki', Chinese characters with level tone were used to transcribe Western Old Japanese low pitch and Chinese characters with other tones to transcribe Western Old Japanese high pitch, but it appears to be true only statistically.

Lexicon: Japanese poetic texts did not use words that, at the time, were obvious loanwords until quite late, so poetic texts contain only native vocabulary and nativized loanwords, such as e.g. *uma* 'horse', *umey* 'plum blossom' (both from Old Chinese *mra and *mɨ), and *tera* 'Buddhist temple' (from Old Korean *tyerV). A few loanwords can be seen in prose texts, most from Chinese or Sanskrit (via Chinese intermediary).

Morphology: The description of word structure and word formation in the present article is based on Western Old Japanese, because of its prestigious nature and due to the less complete data from Eastern Old Japanese. Both Western Old Japanese nominal and verbal morphology are significantly different from any later stages of the language, including Middle Japanese (morphological elements and their functions typical only for Western Old Japanese and not found in later stages of the language are indicated in bold typeface).

Nominal morphology: Most nominals (with the exception of some pronouns) in Western Old Japanese have no formal markers, distinguishing them from other parts of speech, e.g. *yama* 'mountain', *pyito* 'person', *yukyi* 'snow', *myidu* 'water', *piy* 'fire', *puta* '2', *towo* '10'.

The following is the list of the most frequent WOJ affixes. Nominal prefixes: honorific *myi-,* intensive *ma-,* dimunitive *wo-* and *kwo-,* locative **sa-.** Plural suffixes: *-ra* (**neutral** plural marker), *-domo, -tati,* **-na.** Case markers: active-**i**, possessive *-ga,* genitive *-no,* genitive-locative **-tu,** dative-locatives *-ni* and **-ra,** accusative-**absolutive** *-wo*, comitative *-to*, ablative **-ywo/-ywori/-yu/-yuri,** directives **-gari** and *pye,* comparative **-no/-ni/nasu/-nosu,** and terminative *-made*. Dimunitive suffixes: **-ra, rø -kwo.**

Some Western Old Japanese pronouns have two stems: unextended, and extended with suffix *-re*, which may be treated as a formal marker of these pronouns, distinguishing them from other parts of speech, and also from other pronouns that do not have the special stem in *-re*. Personal pronouns: *wa-/ware* 1ps and 1pp (rare), **a-/are** 1ps, **na** 1ps, **na-/nare** 2ps and 2pp (rare), [myi]masi 2ps, **si/so-** 3ps. Reflexive pronoun: *ono-/onore*. It is necessary to note that Middle Japanese has a unified system of personal reflexive pronouns (...self), which does not exist in Western Old Japanese or Eastern Old Japanese. Demonstrative pronouns: proximal *ko/kore* 'this', mesial *so/sore* 'that', distal *ka/kare* 'that over there'. Demonstrative pronouns indicating place or direction: proximal: *koko, koti, konata* 'here'; mesial: *soko* 'there'; distal: *kanata* 'over there'. The distal forms in Western Old Japanese and Eastern Old Japanese may be an innovation, as they are rather rare. In Western Old Japanese there is another distal demonstrative pronoun **woti/woto/wote** that is used more frequently than *ka/kare*, but its usage is predominantly limited to indicating distal place or

time. Interrogative pronouns: *ta-/tare* 'who', *nani* 'what', *ika* 'how', **idu/iduku** 'where', *iduti* 'where to', *idure* 'which', *itu* 'when', **iku/ikuda/ikura** 'how many', *nado/nazo* 'why'. Collective pronouns: *miyna*, **mor̯o** 'all'.

The numerical system of Western Old Japanese sharply contrasts with later stages of the language; numerals of Chinese origin are not yet used, or at least they are not present in the texts. The main problem with attestation of numerals is that they are frequently written logographically, and the phonetic attestations are lacking. The following cardinal numerals are attested phonetically: *pyit̯o* '1', *puta* '2', *myi* '3', *y̯o* '4', *itu* '5', *mu* '6', *nana* '7', *ya* '8', *k̯ok̯on̯o* '9', *t̯owo* '10'. After *t̯owo* '10', attested tens include *pata* '20', *myi-swo* '30', *y̯o-swo* '40', and *ya-swo* '80'. There is only one phonetic attestation of a numeral including both tens and digits: *myiswo-ti amar-i puta-tu* '32' (30-class exceedinf 2-class) that, alongside later glosses to numerals written logographically in Western Old Japanese texts, provides us with the way in which the digits were added to tens: tens + classifier + infinitive of the verb *amar-* 'to exceed' + digit + classifier. Among higher numerals, only the following are attested phonetically: *mwomwo* '100', **i-pwo** '500', **ya-pwo** '800', *ti* '1,000', *y̯or̯odu* '10,000'. The system of classifiers is in its infancy, and only the following classifiers are attested: *-tu* (objects used with digits), *-ti* (objects used with tens and hundreds), *-ri* (people), **-mot̯o** (grassy plants), *-pye* (layers and folds), *-ka* (days).

Verbal morphology: One of the greatest differences between Western Old Japanese and all later stages of the language is that quality verbs like modern Japanese *siro-i* 'white' or Middle Japanese *siro-ki* 'id' are still in the process of formation. This is because in Western Old Japanese stems of quality verbs can still behave as regular adjectives: e.g. <u>*sira* namyi</u> '<u>white</u> waves', <u>*opo* kapa</u> '<u>big</u> river', cf. Middle Japanese <u>*siro-ki* nami</u> 'white waves' and <u>*opo-ki* kafa</u> 'big river' with the obligatory attributive suffix *-ki*, which is not obligatory in Western Old Japanese. Therefore, Western Old Japanese adjectives are in transition from nominals to verbs.

Western Old Japanese verbs are divided into several classes: consonantal verbs, with roots ending in a consonant (*yuk-* 'go', *kyir-* 'cut', *omop-* 'love, think', *n̯ok̯os-* 'leave', etc.), vowel verbs, with roots ending in a vowel (*kwopiy-* 'love', *tasukey-* 'save', *myi-* 'see', *miy-* 'turn', etc.), and irregular verbs (*k̯o-* 'come', *se-* 'do', *ar-* 'exist', *sin-* 'die', *in-* 'go [away]'). In addition, there are defective verbs—*n-*, *t̯o* 'be', *t̯o* 'say'— that have only a limited number of forms and are irregular as well.

Western Old Japanese verbs can take both prefixes and suffixes. The prefixal position is limited to one slot, whereas suffixal positions are multiple. There are a sizeable number of verbal prefixes, most of which occur only in Old Japanese—*i-, ka-, ari-, ta-*, etc.— with mostly unclear or poorly investigated functions. Verbal suffixes can be divided into word-final and word-nonfinal, the main difference between these two groups being that the second group cannot conclude the verbal form and is always followed by another suffix. The rules governing the combination of verbal roots and suffixes are quite complex. However, a general rule of thumb is that suffixes with initial vowel keep the vowel after consonantal roots, and lose it after vowel roots, while consonant-initial suffixes keep initial consonant after vowel roots, but lose it after consonantal roots. Since some vowel-initial suffixes do not lose their initial vowels after vowel roots, but rather cause the last vowel of the verbal root to be dropped, it is convenient to subdivide vowel verbs into strong vowel verbs (that never lose their final vowels) and weak vowel verbs (that lose their final vowels before certain suffixes). Due to these complexities, it is necessary to list verbal affixes in two forms: after consonantal stems and after vowel stems (with the possible differentiation between forms found after weak and strong vowel verbs). Thus, for example the attributive suffix has the following forms: *-u* (<*-uru) after consonantal verb, *-uru* after weak vowel verb, and *-ru* after strong vowel verb: *yuk-u* pyit̯o 'person *who goes*', *kwop-uru* pyit̯o 'person *who loves*', and *myi-ru* pyit̯o "person *who sees*".

It is impossible to provide a list of all verbal affixes here, so only the most important are listed: infinitive *-i/-Ø*, negative infinitive *-azu/-zu,* finite *-u,* attributive *-u/-uru/-ru,* imperative *-ye/-y̯o,* negative *-an-/-az-/-n-/-z-,* tentative *-am-/-m-,* volunative **-ana/-na,** iterative **-ap,** passive **-ye-/?-raye-,** causative **-asimey-/-simey-,** debitive *-ubey-,* negative debitive **-umasizi-,** nominalizer **-aku,** honorific **-as-.**

Besides suffixation, Western Old Japanese widely uses some auxiliaries after the infinitives: **past tense** *-kyi,* retrospective *-kyer-,* perfectives *-te-/-t-* and *-n-,* perfective-progressive *-tar-.* In the later stages of language, these grammaticalized auxiliaries became inseparable from the infinitive marker *-i,* and thus formed new secondary suffixes.

Syntax. Old Japanese syntax remains the most poorly studied area of Old Japanese grammar. Nevertheless, it appears that there were some significant differences with later periods of the language, particularly concerning embedded clauses.

References

Bentley, John R. 2001. A descriptive grammar of Early Old Japanese prose. Leiden, Boston, Köln: Brill.

Hendriks, Peter. 1992. Adverbial and adnominal modification in Old Japanese: copular particles and predications. New Haven, CT: Yale University Ph.D. Dissertation.

Martin, Samuel E. 1987. The Japanese language through time. New Haven, London: Yale University Press.

Miyake, Marc H. 2003. Old Japanese: a phonetic reconstruction. London: Routledge Curzon.

Takeuchi, Lone. 1999. The structure and history of Japanese: from Yamatokotoba to Nihongo. London, New York: Longman.

Vovin, Alexander. 1997. On the syntactic typology of Old Japanese. Journal of East Asian Linguistics 6(3). 273–90.

——— 2005. A descriptive and comparative grammar of Western Old Japanese, Vol. 1. London: Global Oriental.

Whitman, John B. 1985. The phonological basis for the comparison of Japanese and Korean. New Haven, CT: Yale University Ph.D. Dissertation.

ALEXANDER VOVIN

Old Norse

Old Norse (ON), in the widest sense of the term, refers to the varieties of the North Germanic family used in Scandinavia and its colonies from the breakup of its parent Germanic until around the fourteenth century. A narrower usage would restrict ON to Old Norwegian (ONo) and Old Icelandic (OIc) between the seventh and fifteenth centuries; still more restrictively, Classical ON (ClON) refers to the literary language of Iceland from around 1150 to 1350, a period well attested in saga literature, and it is this variety from which examples will generally be drawn.

As a member of the Germanic branch of the Indo-European family, ON preserves a number of archaic features, which make it useful for comparison with other early Indo-European languages. For nouns, two numbers, three genders, and four cases are distinguished; around 14 declension patterns occur, based on original stem type. Adjectives and pronouns agree with nouns in number, gender, and case. Many verbs retain the ancient vowel ablaut to mark tense, e.g. bita—beit—bitinn 'to bite—bit (sg.)—bitten' (compare Greek leipo—leloipa—elipon). From the Proto-Germanic (PrGmc) period, evidence of the Grimm's Law consonantal shift remains, e.g. fiskr 'fish', kinn 'cheek' (compare Latin piscis, gena), and more indirectly, of Verner's Law, e.g. finna—fundu 'to find—they found', vas—varu 'was—were'. New dual first and second person pronouns, a double declension of adjectives, and a past tense suffix containing a dental consonant are also PrGmc innovations retained in ON.

The earliest period of Norse history covers the gradual separation of the language from PrGmc, approximately between the first and sixth centuries CE. This era is often termed urnordische/urnordisk (Proto-Norse/Ancient Scandinavian, etc.), although since many changes were still shared with WGmc (ancestor of Old High German (OHG), Old English (OE), Old Saxon (OS), etc.) but not with EGmc (i.e. Gothic (Got)),

it might be regarded as Proto-North-West Germanic rather than specifically Norse. Thus, PrGmc *ae became NGmc and WGmc *a but EGmc *e e.g. ON mani OS mano vs. Got mena 'moon'; likewise, PrGmc *z developed into NGmc and WGmc *r (in NGmc, perhaps via some kind of palatalized fricative, usually represented as R) but remained in EGmc (or was devoiced,) e.g. ON eyra OHG ora vs. Got auso 'ear'. However, some have regarded parallels between NGmc and EGmc as more significant for grouping the languages, e.g. PrGmc *ww gave NGmc and EGmc *ggw but WGmc *uw (then becoming a diphthong), e.g. ON tryggr (acc. tryggvan) Got triggws vs. OE triewe 'true'.

The 'Norse' of this period is attested in a number of inscriptions in the runic alphabet (the older futhark, containing 24 runes), some of which predate the fourth-century Gothic biblical translation, and are thus the earliest direct evidence for Germanic. They are found mainly in Denmark and southern Norway and Sweden, and reflect a time when the phonology and morphology of PrGmc were still well preserved. Thus, the Gallehus horn inscription from Denmark from around 400 CE reads ek hlewagastiR holtijaR horna tawido—'I, Hlegest son of Holti, made (the) horn', which would produce the shorter ClON equivalent ek Hlegestr Holtir horn tadha.

Between roughly the sixth and tenth centuries, various important changes further distinguished Norse from the rest of Germanic, while internal distinctions were minimal. This period might be termed Common or Viking Norse, since from around 800 to 1000 viking expansions carried Norse speakers outside Scandinavia itself, sometimes resulting in lasting settlements: from Norway, west to Scotland, Ireland, the Faroes, Iceland, and Greenland; from Denmark, to England, northern France, and eastern Prussia; from Sweden, east to Gotland and Russia. The language is still mainly known from runic inscriptions, largely in the younger 16-rune futhark, although much poetry preserved in

later manuscripts, in the court (skaldic) and heroic (eddaic) traditions, was composed orally in this period.

During Common Norse (CN), unstressed vowels were reduced to only three qualities, ceased to distinguish length, and were completely lost in many phonetic environments. This caused the loss of a number of morphological markers and hence several mergers and analogical reformations in the nominal system. However, many distinctions were retained through a shift of information from (often lost) final vowels to stressed root vowels: by the process of mutation or umlaut, root vowels underwent partial assimilation to following sounds, and these new pronunciations were then phonologized, expanding the stressed vowel inventory. Thus, through palatal mutation, a following unstressed [i] (or [j]) changed [a] to [E] (later written <ae> or <e>), [o] to [o] (written <oe> when long) and [u] to [y] (i.e. fronting/raising), while labial mutation (by [u] and [w]) rounded [a] to [o] (written <o>), [i] to [y], and new [E] to [oe] (written <o>), i.e. five new vowel qualities were created. A following unstressed [a] also caused lowering of preceding high vowels, and breaking/diphthongization of preceding [e] to [ia]. Morphological alternations such as fotr—foetr (from *fotiR) 'foot—feet', land—lond (from *landu) 'land—lands', and geldr—gjalda (from *geldan) 'pays—to pay' were thus established.

An apparent irregularity in the operation of i-mutation has attracted considerable interest among Norse researchers. The vowel of a short syllable does not seem to have been mutated by lost [i], e.g. katlar (from *katiloR) 'kettles' as opposed to the mutated singular ketill (from *katilaR). The traditional theory of Axel Kock explains this as the result of i-mutation having taken place in two different periods: first, only long syllables were affected, and the following [i] was then lost; later [i] was generally lost after short syllables; finally, all instances of remaining [i] caused mutation. However, this model does not appear entirely satisfactory, and a range of other explanations have been presented.

Several notable grammatical developments also occurred during CN. A new set of verbal forms with reflexive/medio-passive significance arose from the suffixation of the reflexive pronoun sik to the verb, e.g. berja 'to beat, strike' vs. berjask 'to fight'. The combination of 'to have' with the neuter past participle produced new perfect and pluperfect constructions, e.g. fadhir hans hafdhi haldit skottum fyrir Hakoni 'his father had withheld taxes from Hakon'. A definite article was developed from the demonstrative hinn, which could be suffixed to nouns, e.g. medh ollum farminum 'with all the cargo'; in arma jotna systir 'the wretched giants' sister'. A pitch accent distinguishing monosyllables and bisyllables may also have come into existence at this time, although it was not shown in writing.

From the tenth century onward, dialectal differences began to develop between West Norse (WN), spoken in Norway and its colonies (chiefly Iceland), and East Norse (EN), spoken in Sweden, Denmark and their colonies (although Gutnish might be regarded as intermediate between WN and EN). A notable distinction was the EN monophthongization of CN [Ei] > [e:] and [ou], [Ey] > [o:], e.g. Old Danish (ODa) enn 'one', dothaer 'dead' vs. Old Icelandic (OIc) einn, daudhr. Mutation by [i] and [u] was carried out more completely in WN than EN, and palatal mutation was also caused by -R and by [g]/[k] before high front vowels in WN, e.g. OIc nofnum 'names (dat.)', thaer 'they (fem.)' vs. Old Swedish (OSw) nafnum, thar. By contrast, breaking occurred more widely in EN.

During the late tenth and early eleventh centuries, the Scandinavian countries were converted to Christianity, and Norse began to be written in a modified form of the Roman alphabet (borrowing the rune thorn <th> and the letter eth <dh> from Old English usage in some regions to represent the dental fricatives, and adding diacritics to some vowels). The oldest surviving manuscripts in Norse are, however, from the late twelfth century. From this era until around 1350 (when the Middle periods of the Scandinavian languages might be said to begin), dialectal distinctions within EN and WN became increasingly marked, and one can refer to OIc, ONo, etc.

EN, and in particular ODa, was generally more innovative than WN. ODa merged the unstressed vowels as schwa (written <ae> or <e>), e.g. ODa brothaer 'brother', dottaerson 'daughter's son' vs. OIc brodhir, dottur-son. Intervocalic consonants underwent lenition, with voiceless stops being voiced, voiced stops sometimes becoming fricatives, and voiced fricatives becoming semivowels or being lost, e.g. ODa nogaer 'some, a certain', sauthe 'said' vs. OIc nokkurr, sagdhi. The case system was substantially reduced, and there was leveling among the personal endings of verbs. The pitch accent distinguishing original monosyllables from bisyllables was changed into a reduced glottal stop, termed the stod, in stressed syllables with long vowels or final voiced consonants; this accent became lexically distinctive where old monosyllables became bisyllabic through the insertion of epenthetic vowels to break up consonant clusters, or through suffixation of the definite article.

OSw underwent many of the same developments as ODa, but to a lesser degree, and later. A phenomenon of vowel balance affected unstressed vowels: after long syllables, they were pronounced more weakly, being written as <ae e o> in some manuscripts, and eventually merged and underwent apocope; after short syllables, however, they retained their qualities longer, being written as <a i u>, and were more resistant to reduction and loss, e.g. mother 'mother' vs. fathir 'father', gangae 'to go, walk' vs. fara 'to go, travel'.

ONo, while generally similar to OIc, shared some developments with EN, e.g. initial clusters of [hl] and [hr] were simplified to [l] and [r]: ONo lut 'share (acc.)', ratt 'pushed' vs. OIc hlut, hratt. Some parts of ONo exhibited vowel harmony of unstressed vowels: following nonhigh vowels, [i u] were lowered to [e o], e.g. honom 'him (dat.)' vs. sinum 'their (dat.)', kononge 'king (dat.)' vs. bili 'moment (dat.)'.

OIc (i.e. ClON) was generally a very conservative variety; however, it innovated in some ways. Around the start of the literary OIc period, the language had a seven-vowel basic 'triangle': three front vowels [i e E] and four back vowels [u o o a], with two front rounded vowels [y o] that resulted from the CN mutations; all could be either long or short, and nasalized long vowels were also contrastive. There were three falling diphthongs: [Ei ou Ey]. This system is described in the remarkable twelfth-century native work known as the First Grammatical Treatise. During the literary period, the low–mid vowels underwent mergers: [E] merged with [e], [E:] with [o:], [o] with [o] (later written <o>), and [o:] with [a:], e.g. OIc segidh 'say!', fell 'fell' vs. ONo saeghit, fell. Nasalization was also lost; [y] merged with [i]; [e:] diphthongized to [jE] and [o:] to [ai].

The consonantal system of ClON underwent limited change from CN. Before [l], [r] and most rounded vowels, [w] was lost; it survived as [v] in some environments, e.g. OIc reka 'to drive', ordhinn 'having become', vidh 'against' vs. OSw vraekae, wordhin, Old Gutnish withr. Velar stops palatalized before front vowels and [j] during the literary period, and consonant clusters ending in sonorants were broken up with epenthetic [u], e.g. Modern Icelandic madhur 'man' vs. OIc madhr. If a tonal accent existed in CN, it was lost in Icelandic.

Grammatically, the morphology of CN was very well preserved in ClON. There were a number of loanwords from Latin, English, or German associated with the new Church, e.g. kirkja 'church', paskar 'Easter', biskup 'bishop'. Word order was fairly free, although with a tendency to subject–verb–object, in prose at least. A notable feature of narrative style was the frequent use of the historical present, with tense often switching casually within sentences, e.g. Sidhan logdhu their saman flotann; teksk thar in grimmasta orrosta ok fell mart af hvarumtveggjum… 'Then they brought the fleets together; the fiercest battle begins there and many fell on both (sides)…'.

ON is worthy of study not only for its intrinsic linguistic interest, but also for the access it gives to a wealth of literature, reflecting a time when pagan and Christian worlds met, when explorers and conquerors expanded out of their original homelands, and when the modern nations of Scandinavia were born.

References

Antonsen, Elmer. 1975. A concise grammar of the Older Runic inscriptions. Tubingen: Max Niemeyer.

Benediktsson, Hreinn. 1982. Nordic umlaut and breaking: thirty years of research. Nordic Journal of Linguistics 5.

Faarlund, Jan. 1994. Old and Middle Scandinavian. The Germanic languages, ed. by Ekkehard Konig and Johan van der Auwera. London, New York: Routledge.

Gordon, Eric. 1927. An introduction to Old Norse, revised by Arnold Taylor. Oxford, New York: Oxford University Press; 2nd revised edition, 1957.

Haugen, Einar. 1950. The first grammatical treatise: the earliest Germanic phonology. An edition, translation and commentary. Baltimore: Linguistic Society of America; 2nd revised edition, London: Longman, 1972.

———. 1976. The Scandinavian languages: an introduction to their history. London: Faber and Faber.

Kristjansson, Jonas. 1988. Eddas and sagas. Reykjavik: Hidh islenska bokmenntafelag; 3rd edition, 1997.

Nielsen, Hans. 1979. De germanske sprog: baggrund og gruppering. Odense: Odense universitetsforlag; as The Germanic languages: origins and early dialectal interrelations. Tuscaloosa, London: University of Alabama Press, 1989.

Noreen, Adolf. 1884. Altislandische und altnorwegische Grammatik (Laut- und Flexionslehre) unter berucksichtigung des urnordischen (Old Icelandic and Old Norwegian grammar (phonology and morphology) in view of Old Norse). Halle (Saale): Max Niemeyer; 4th revised edition, 1923.

———. 1904. Altschwedische grammatik, mit einschluss des altgutnischen (Old Swedish grammar, with the inclusion of Old Gutnish). Halle: Max Niemeyer.

Riad, Tomas. 1998. The origin of Scandinavian tone accents, Diachronica 15.

Skautrup, Peter. 1944. Det danske sprogs historie (History of the Danish language), Vol. 1. Copenhagen: Gyldendal.

MARISA LOHR

See also **Gothic; Old High German; Old English; Swedish and Scandinavian Languages**

Old Tibetan

Old Tibetan is the ancestor of all modern Tibetan dialects. This language is attested from the seventh century onward. It was originally spoken in a relatively tiny area (if we compare with historical Tibet's huge size). The origin of the Tibetan people was most probably the Lhoka in southern Tibet, where remain the Yum-bung

bla-sgang, the oldest fortress in Tibet, and the royal tombs in 'Phyong-rgyas. From Emperor Srong-btsan sGam-po at the beginning of the seventh century up to the middle of the ninth century, the Tibetan Empire became one of the major powers in Central Asia. After this brief period of expansion, the Tibetan-speaking area has not changed much up to now.

Old Tibetan should not be confused with Classical Tibetan, which is still the standard written language for most Tibetans. Classical Tibetan is an old dialect that is not the direct ancestor of all modern dialects, but it is the basis for the orthography of Modern Literary Tibetan, also called 'Written Tibetan'. We consider only texts older than the tenth century as genuine 'Old Tibetan'.

The Buddhist conversion of Tibet entailed the coinage of new terms, and the creation of a dictionary, the Mahatvutpatti (*bye brag rtog ched*) at the beginning of the eighth century. Besides, the orthography underwent several reforms by the end of the Imperial Period. Few documents exist in genuine Old Tibetan. There are the famous Sino-Tibetan bilingual treaties inscribed on pillars (*rdo ring*), the extensive Dunhuang documents that have been partially translated into French, English, and Chinese, and only two transmitted texts that preserve the most archaic features: *the bKa-chems ka-khol-ma* (testament of Srong-btsan sGam-po) and the *sBa-bzhed* (chronicle of bSam-yas). Many texts from the Buddhist canonical scriptures (*bKa-'gyur* and *bsTan-'gyur*) date from the Imperial Period, but they were rewritten so many times that most archaic features have been weeded out. No grammar of Old Tibetan exists yet and the extent of the difference between the Classical and the Old language is not fully understood.

Language Classification

Tibetan is usually classified as a Bodic language. The Tamang-Gurung-Thakali languages of Nepal are generally considered its closest relatives. However, no reconstruction of the ancestor language Proto-Tibetan based on the comparison with these languages has been undertaken so far; hence, our knowledge of the prehistory of Tibetan is mostly based on internal reconstruction and on comparison with more remote cousins like Archaic Chinese and Burmese. Some scholars have proposed that Tibetan is especially close to Chinese within the Tibeto-Burman family, but further investigations are needed.

Vocabulary

Tibetan has few loanwords from either Sanskrit or Chinese. Indic words are limited to animals and plants not found in Tibet (*seng-ge* 'lion', *pad-ma* 'lotus', *ping kyur ma* 'a kind of bird'). Loanwords from Middle Chinese are restricted to cultural vocabulary (*lcog tse* 'table', *dong tse* 'coin'). The loanwords from Archaic Chinese have never been studied yet, as it is difficult to tell apart loanwords from words with the same ancestry (cognates) (one such example is *lcags* 'iron' : it must be a loanword, since iron appeared only in the fifth century BC in China). Besides, Tibetan absorbed many non-Tibetan languages during the expansion of the empire, and these have certainly influenced its vocabulary. This might be one cause for the existence of redundantly duplicated words with similar meaning, such as *rmi* and *rmang* 'dream'.

Tibetan has several striking lexical innovations, such as *bdun* 'seven' (Burmese *khu hnac*), *khyed* 'you' (Burmese *nang*), *khrag* 'blood' (Burmese *sweh*), not found in any other language of the family.

Clear traces of the derivational morphology inherited from Proto-Sino-Tibetan are found in Tibetan. However, it probably was not in active use anymore even by the time of the first Buddhist translations, since new means were created to coin new terms: instead of using the -s suffix deriving an action noun from a verb (*za* 'to eat' *zas* 'food'), translators chose to add the verb *byed* 'to do' in its future form *bya* after a verb to translate the Sanskrit past participle.

Writing System and Phonology

The phonology of Old Tibetan is obscured by the irregularities in spellings found in most Dunhuang documents, which often make the translation quite difficult if not impossible. The exact pronunciation of Tibetan in the imperial period is mostly uncontroversial, except for a few letters. The prefixed '-, incorrectly called 'a chung' even by Tibetan specialists ('a chung designates in fact the little 'a used to indicate vowel length in Sanskrit words), has been the object of some debate, as it seems to have three distinct functions in Tibetan orthography: when used as an initial, it represents a voiced glottal fricative (as in *'od* 'light'). This sound is still heard in some dialects. As a preinitial, it indicates prenasalization (as in *'khor* 'to turn', most likely pronounced as *ηkhor*). Finally, since the vowel a is not written by any special symbol in the indic alphabet, this letter also serves to disambiguate some combinations of letters (*mda'* 'arrow', written M D ' as opposed to *mad* 'under' written M D).

The final stops -b -d -g are written with the symbols for voiced consonants, although these were most probably pronounced as unreleased stops as in most languages in the area.

Old Tibetan boasts of a great quantity of complex consonant clusters that are preserved in modern orthography but only partially in dialects. These clusters are extremely interesting typologically, since

some of them are highly unusual: clusters such as *rt-* are permitted, but not the cross-linguistically much more common *tr-*.

Tibetan admits up to four initial consonants (*brgyad* 'eight'), but no more than two finals (*dmangs* 'people'). One of the most conspicuous differences between Old and Classical Tibetan is the *da drag*: while in Classical only -s can stand as a second final, in Old Tibetan some clusters exist with -d as a second final, such as in *mnard* 'he oppressed' (Classical *mnar*).

In the Tibetan writing system, there are three series of initial consonants : voiced, aspirated, and plain unvoiced. The former can appear in all contexts, but the latter two seem to be variants of a single type in Classical Tibetan: except for some rare exceptions, only aspirated consonants can begin a word or stand after nasal preinitials (m- and '-), while when preceded by any other preinitial (stop, liquid or s-), only plain voiceless stops are permitted. This reminds one of English stops (p is aspirated in *pit* and nonaspirated in *spit*). In Old Tibetan, the distribution of these graphemes is much less predictable (*gchig* systematically for Classical *gcig* 'one', for instance), but it is not probable that these two series are truly distinctive in Old Tibetan.

Finally, there is a vowel in Old Tibetan never found in the Classical language: the inverted I, transliterated as *ï*. This sign does not seem to represent a phoneme distinct from *i*, though. The rules governing their distribution are still unknown.

Verbal System

Like modern Tibetan dialects, Old Tibetan was an ergative language, which means that the object of a transitive verb bears the same mark as the subject of an intransitive verb (in Old Tibetan, this case has no overt marking), while the subject of a transitive verb is marked with ergative case (variously marked as 'is, kyis, gyis, or gis).

klu-gong glo-ba nye'o (Zhol inscription, South:20)
Klu-gong (no suffix)—intention—near—finite verb ending
Klu-gong was loyal.

gzhan '**is** **bdag** bsad **sha** zos dka' myi dka' (Pelliot Tibétain 126: 20)

other—ergative suffix—**yourself** (no
suffix)—kill—**flesh** (no
suffix)—eat—hard—not—hard
If others killed you and ate (your) flesh, wouldn't it be hard?

The ergative suffix, however, serves as an intrumental case and can mark emphasis.

Unlike Kiranti or rGyalrongic languages, Tibetan has no trace of agreement between verb and e.g. the subject. Transitive verbs have four forms (present, past, future, and imperative), and intransitive verbs have two forms. In most cases, one form is not sufficient to derive the four forms, so that the verbal conjugation is far from simple. Here are examples of verbal patterns found in Old Tibetan (for some forms unattested in Old Tibetan, we provide Classical Tibetan in italics):

Present	Past	Future	Imperative	
byed	byas	bya	byos	
'dzin / zin	bzung	gzung	*zung*	to take
'dogs	btags	gdags	*thogs*	to tie
za	zos	bza	zos	to eat
sbyor	sbyard / spyard	sbyar	*sbyor*	to join

Some elements of the verbal conjugation are quite old: the past tense -s suffix is also found in rGyalrong, Qiang, and Kachin, and the change of a to o in the imperative is related to the imperative -o suffix found in Tamang. However, many peculiarities of the verbal system (the change a to o in the present tense of some verbs, the past tense b- prefix) do not seem to be relatable to anything in other Tibeto-Burman languages.

An interesting irregularity in conjugation that is common to Tibetan and Kiranti languages is the conjugation of the verb 'to eat'. In Tibetan, it presents an exceptional vowel alternation in the past tense a à o unrelated to the more common alternations found in the present and imperative. In Hayu, 'to eat' is one of the four verbs with a -o stem alternation: *dza-ng* (I eat it), *dzo* (You eat it). The same irregularity is found in a dialect of Limbu. This must be an archaism going back far in the prehistory of the language.

References

Bacot, J.F.W. Thomas, and C. Toussaint. 1940–1946. Documents de Touen-houang relatifs à l'histoire du Tibet. Paris.

Huang, Bufan, and De Ma. 2000. Translation and annotations on Dunhuang Tibetan manuscripts about the history of Tufan period. Beijing.

Imaeda, Y. and T. Takeuchi. 1990. Choix de documents Tibétains, tome III: corpus syllabique. Paris.

Imaeda, Y. et al. 2001. Choix de documents Tibétains, Tome IV: corpus syllabique. Tokyo.

Li Fang Kuei, and W. South Coblin. 1987. A study in the Old Tibetan inscriptions. Taipei.

McDonald, A. 1971. Une lecture des Pelliot Tibétain 128/6, 1287, 1038, 1047 et 1290. Etudes tibétains dédiées à la mémoire de Marcelle Lalou, 190–391. Paris.

Richardson, H.E. 1985. A corpus of early Tibetan inscriptions. London.

JACQUES GUILLAUME

See also **Sino-Tibetan Languages**

Onomatopoeia

Onomatopoeia is better understood as a semantic phenomenon if language is likened as a diagram. A referent is codified through the phonetic-symbolic means of language into a word, i.e. the linguistic depiction of its representation. All languages show some kind of imitative linguistic codification known as onomatopoeia. Onomatopoeia represents a referent based on a diagrammatic sound–meaning identification. This direct sound–meaning relationship is nonarbitrary or isomorphic. Examples are easily found in many languages, especially the use of animal sounds like the Malay *kokok* 'the crow of a rooster', *embek* 'the bleat of a goat', the Mandarin *gaga* 'the quack of a duck', and *miao miao* 'the meow of a cat'. Onomatopoeia as a universal language feature can be divided into three types: sound imitation, secondary sound symbolism, and echo.

Onomatopoeia of the first type occurs when linguistic sound imitates sounds of the real world. Each linguistic imitation is taken as the diagrammatic sign of the signified referent. Many terms are derived from sound imitation, for example, *cuckoo, cockatoo, boom, buzz, rumble, twang, chirp, slam, thump, clap, flap,* and *rattle*. More common onomatopoeic words like *ha ha ha* that signifies laughter in English and Cantonese opera is used as a sarcastic tag as in 'How funny, ha ha ha' in specific contexts of communication.

Onomatopoeic words are salient in Malay vocabulary as illustrated in Table 1.

Sound symbolism occurs when there is a recurrence of sound–meaning correlation. An example would be the high front unrounded vowel taken to symbolize a sense of small, insignificant, slight, or weak. The vowel contrast in *pinch* vs. *punch*, *gripe* vs. *grope*, and *sip* vs. *suck* points to this sound–meaning symbolism in /i/. Research shows that the distribution of sounds in men's and women's names is not random. It is common for Caucasian women's names to end with /i/ and men's with consonants like fricative, i.e. /s, f/, and stop, i.e. /t, k/. The psychology experiment concludes that the gendered sound difference in personal names is best understood as reflecting emotional differences. Men's names are associated with active, cheerful, nasty, and unpleasant emotions while women's names are associated with passive, sad, soft, and pleasant emotions (see Whissell 2001). If language is diagrammatic, personal names are depictions of human beings whose respective names become the aspirations intended for them.

Secondary sound symbolism in the onomatopoeia context is a linguistic sound taken to connote another meaning based on phonetic similarity. Mandarin and Cantonese speakers believe that the sound of 8, i.e. *ba,* is sound symbolic to *fa* 'prosper'. As such, car registration numbers that contain many 8's are more expensive in Malaysia and Singapore. Since 4 *sì* is sound symbolically related to *sì* 'die' in these languages, karaokes in Singapore and Malaysia will add lucky phrases to tags with the number 4 before offering this queue number to customers. One can imagine how confident a typical Chinese needs to be in order to drive in a car with 4444 as the registration number.

Secondary sound symbolism plays an important role in Chinese speech communities when one chooses a name for a child. *Lóng* 'dragon', a symbol of divinity and good fortune, is a favourite character used for boys. An excellent example would be *Chen Long* (Jackie Chan), who became the dragon of Chinese cinematography. Bruce Lee's middle name is also the character of a dragon. Other favorite characters used as Chinese personal names are *fa* 'prosperous' as in *Chow Yuen Fatt*, *an* 'safe' as in Ang Lee, and *kuang* 'bright' the first name of the former Singapore Prime Minister. Onomatopoeia in the form of secondary sound symbolism is culturally significant as it influences one's belief.

TABLE 1

Onomatopoeic Lexicon	Meaning
Debuk	The sound of slapping or boxing
Debum	The sound of heavy things falling
Debung	The sound of drumming
Debur	The sound of waves breaking on the shore
Debus	The sound of flapping of wings
Debut	The sound of squeezing out of a small hole
Debak	The sound of boxing each other
Debar	The sound of heart pounding
Debik	The sound of hitting with one's palm
Degar	The sound of blowing with a stick
Deguk	The sound of gulping water
Dekah	The sound of hearty laughter
Deham	The sound of a little cough
Dengkung	The wailing of a dog
Dengung	The sound of a plane
Dengkur	Snoring
Dentang	The sound of hitting the iron
Denting	The sound of falling coins
Dentung	The sound of thunder
Dering	The ringing of the phone

TABLE 2

Semantic Criterion	Examples
Motion	huff-puff, flip-flap, flipperty-flopperty
Tumult	rumble-jumble, hubble-shubble, wringle-wrangle
Inferiority	shally-wally, phoney-baloney, whimsy-whamsy
Trickery	hooky-crooky, hanky-panky, cuddle-muddle
Mental instability	randy-dandy, shilly-shally, hazy-dazy
Mutuality	equal-aqual, hob-nob, tiggle-taggle
High degree	tip-top, wee-jee, itsy-bitsy
Hypocoristic	mousey-pousey, tootsy-wootsy, lovey-dovey

Echo is a form of reduplication that occurs in the form of freezes, binomials, echo formation, and other forms of bipartites. Echoic morphology is language universal. Table 2 illustrates some English echos.

Malay echoic reduplications like *gunung-ganang* 'ranges of moutain', *batu-batan* 'rocks', *sayur-mayur* 'all kinds of vegetables', *bengkang-bengkok* 'very crooked', and *cucu-cicit* 'descendants' are commonly used in daily conversation. Malay partial reduplication is important in word fomation, especially in naming things: *tikus* 'mouse' > *tetikus* 'computer mouse', *kunci* 'key' > *papan kekunci* 'keyboard', and *pejal* 'solid (adjective)' > *pepejal* 'solid (noun)'. South Asian echos, on the other hand, are loaded with pragmatic functions. Bhojpuri echos like *deslai-oslai* 'matchbox etc.' is used as an indirect request for cigarettes by youngsters or subordinates because smoking in the presence of parents or elderly persons is bad manners. In Tamil, *vendu* 'come' is echoed as *vendu gindu* 'come etc.'. The echo is used for the addressee to come (and run an errand for him). Cantonese echos like *yùhn lūk lūk* 'rounded', *chī lahp lahp* 'sticky/goey', *dung bīng bīng* 'freezing cold', and *baahk syūt syūt* 'snow-white' come in tripartites. These echos illustrate more intense

meaning, i.e. quantity iconicity where more forms entail more meaning.

Despite the skepticism about the relevance of onomatopoeia, it remains resilient in language. The fact that onomatopoeia exists through time and manifests creatively in various forms testifies to the significance of this component either as pragmatic tool, cultural routine, or semantic paradigm that prevails in language use.

References

Abbi, Anvita. 1992. Reduplication in South Asian languages: an areal, typological and historical study. New Delhi: Allied Publishers.

Bolinger, Dwight. 1992. Sound symbolism. International Encyclopedia of Linguistics 4.

Haiman, John. 1985. Natural syntax: iconicity and erosion. Cambridge: Cambridge University Press.

Hinton, Leanne, Johanna Nichols, and John Ohala (ed.) 1994. Sound symbolism. Cambridge: Cambridge University Press.

Jakobson, Roman, and Linda R.Waugh. 1979, 1987. The sound shape of language. Berlin: Mouton De Gruyter.

Jespersen, Otto. 1922. Symbolic value of the vowel i. Philologica 1.

Sew, Jyh Wee. 1995. Sound meanings in Malay. M.A. Thesis. Singapore: The National University of Singapore.

———. 1996. Symbolisation in Malay: evidence in genres and lexicon. Proceedings of the 4th International Pan-Asiatic Symposium of Language and Linguistics, Vol. I. Salaya: Mahidol University Press.

Thun, Nils. 1963. Reduplicative words in English: a study of formations of the types tick-tick, hurly-burly and shilly-shally. Sweden: Carl Bloms.

Tiwary, K.M. 1968. The echo-word construction in Bhojpuri. Anthropological Linguistics 10.

Waugh, Linda R. 1993. Against arbitrariness: imitation and motivation revived, with consequences for textual meaning. Diacritics 23.

Whissell, Cynthia. 2001. Sound and emotion in given names, Names 49.

JYH WEE SEW

See also **Iconicity**

Oto-Manguean Languages

Oto-Manguean languages are distributed throughout Mesoamerica and form the language stock with the deepest time depth in the Americas. The emergence of its eight linguistic families patterns geographically with the ecological zones of the domestication of maize, chili, squash, and beans, which emerged as primary

agricultural products among the peoples of Archaic Mexico beginning about 4400 BCE. In this account, early Oto-Manguean diversified on a landscape where smaller, more mobile bands came to rely increasingly on domesticated plants and center their production in villages on the mesas and highland valleys of the

Central Altiplano, the Mixteca Alta, the Valley of Oaxaca, and the Tehuacán Valley. These regions help define four branches of Oto-Manguean, with Otopamean-Chinantecan and Tlapanecan-Manguean forming the Western division and Amuzgo-Mixtecan and Popolocan-Zapotecan the Eastern division.

Most of the language families of West Oto-Manguean are centered in the highlands and watersheds from the Central Altiplano to the Pacific in Guerrero, and East Oto-Manguean families in highlands and watersheds from the Mixteca Alta to the Pacific in Oaxaca. Some consonant differences between the separate reconstructions of the ancestor languages for East and West suggest that the oldest accessible form of Oto-Manguean may already show dialect differentiation, possibly resulting from dispersal among these watersheds. The connections between the Chiapanecan and Manguean languages in the far east of Mesoamerica and Tlapaneco of West Oto-Manguean are accounted for as the result of population migrations in prehistory.

West Oto-Manguean includes the two branches of Oto-Pamean-Chinantecan and Tlapanecan-Manguean. Northern Oto-Pamean consists of Chichimeco of Guanajuato and the Pame languages of San Luis Potosí. Southern Oto-Pamean consists of Matlatzinca and Ocuilteco around the Valley of Toluca, and an Otomian group comprising the Mazahua language and six branches of Otomí in the states of Mexico, Tlaxcala, and Michoacan. The Chinanteco language family represents the extension of the stock east of the Sierra Madre and contains at least six languages. The Tlapanecan-Manguean branch consists of Tlapanecan-Subtiaban and Chiapanecan-Manguean. Tlapaneco is spoken in several varieties in the state of Guerrero, and Subtiaba was historically spoken near the Pacific coast of Nicaragua. Chiapanecan-Manguean consisted of Chiapaneco in Chiapas and Mangue in coastal Honduras and Nicaragua.

East Oto-Manguean includes the Popolocan-Zapotecan and Amuzgo-Mixtecan branches. Popolocan comprises four Mazateco languages of Oaxaca, Puebla, and Veracruz; four Chocho-Popolocan languages of Oaxaca and Puebla; and Ixcateco of Oaxaca. Zapotecan includes three Chatino languages in southwestern Oaxaca and five branches of Zapoteco in the mountains of southern, western, southwestern, and northern Oaxaca, and in the Valley of Oaxaca east to the Isthmus of Tehuantepec. Each branch of Zapoteco further divides into several distinct languages. Amuzgo-Mixtecan is made up of Amuzgo in Guerrero, two Trique languages in Oaxaca, and Mixteco-Cuicateco, which itself includes the Cuicateco language of Oaxaca and several divergent Mixtecan language areas in Guerrero, Puebla, and Oaxaca.

Over 100 years of Oto-Manguean scholarship provides us with evidence for several statements about the grammatical structure of Proto-Oto-Manguean, the assumed ancestor language. Simple words typically had only one syllable, and complex words could be formed by compounding simple word roots or by attaching prefixes. The language most likely had tone distinctions; i.e. melodic pitch differences between otherwise identical words could alone indicate a meaning difference. The vocalic system included minimally *i, *e, *a, *u, several diphthongs, the semivowels *w and *y, and the consonants *t, *ts, *k, *kʷ, *s, *x, *xʷ, *l, *n, *m (the asterisks indicate that these sounds have been reconstructed on the basis of documented languages).

While parts of the Oto-Manguean family as presented here were separately grouped by researchers of the nineteenth and early twentieth centuries, it was Edward Sapir's 1929 classification that was the first to include the earlier groupings of Mixtecan-Zapotecan, Otomian, Chiapanecan-Manguean, and Chinantecan in a single group. From the 1940s until the mid-1960s, Oto-Manguean research had a blossoming period during which several dedicated scholars worked out the relations of diversity within the Oto-Manguean branches. A student of Sapir, Morris Swadesh, compiled the first comparative study of the sound correspondences of the Zapotecan languages. In the 1950s, Sarah Gudschinsky worked on the internal diversity of Popolocan and supported a larger grouping that included the East Oto-Manguean families of Mixtecan and Popolocan, which she called 'Proto-Popotecan'. Eric Hamp refined and revised both Gudschinsky's earlier subgrouping of Proto-Popolocan and the analysis of the sound system of Chiapanecan-Manguean. María Teresa Fernández de Miranda also worked intensively with the Popolocan language family and the Manguean family and was working out a reconstruction of Proto-Zapotecan before her untimely death. Robert Longacre's extensive work on the internal relationships of Amuzgo-Mixtecan also began with publications in the 1950s. In the same decade, our knowledge of West Oto-Manguean developed with Stanley Newman and Robert Weitlaner's Proto-Otomí reconstructions. In the 1960s, Doris Bartholomew contributed work on Oto-Pamean and revised the Proto-Otomí reconstructions. Along with Michael Piper, she prepared Fernández de Miranda's unfinished Proto-Zapoteco reconstructions for publication.

A virtual dearth of publications spans over a decade before the publication of the most influential work on Proto-Oto-Manguean. Calvin Rensch began systematic work on Proto-Oto-Manguean in the early 1960s. A brief period of new interest developed after the 1976 publication of Rensch's 1966 thesis, *Comparative Oto-*

Manguean phonology, in which he focuses on reconstructing the tone-carrying syllables of words in Proto-Mixtecan. Based largely on Rensch (1976), J. Kathryn Josserand, Marcus Winter, and Nicholas Hopkins's (1984) *Essays in Oto-Manguean culture history* treated the unique opportunity that Mesoamerica provides for integrating historical linguistic and archeological approaches to culture history. In the 1980s, a student of Eric Hamp, Terrence Kaufman, critically assessed the word and sound correspondences contained in the work of Rensch, Longacre, and other historical sources, and in 1994 presented the classification adopted here.

The work of the handful of scholars mentioned above only begins to point to the interest of this language stock for linguistic and historical linguistic theory. The bulk of the above works focus mostly on the sound system and the vocabulary. However, the languages of Oto-Manguean also present challenges for research in intonation, sentence grammar, word structure, and meaning, for which the literature is almost entirely lacking.

The Oto-Manguean sound system exhibits some unique characteristics. For example, many of the languages lack [p] and [b], and most consonants are voiceless, i.e. they are produced, like [t, k, s], without vocal cord vibration. Voiced consonants such as [d, g, z], on the other hand, are rare.

Oto-Manguean is the world's oldest language stock that provides sufficient evidence for tone distinctions in the ancestor language. The complicated tone systems distinguish between two and six pitch levels; i.e. in the most complex scenario, one could theoretically have six different one-syllable words differing only in the pitch or the pitch pattern of the vowel.

'Whistled speech' has been reported in Mazateco, Mixteco, and Zapoteco. This is a form of speech that is basically stripped of individual segments and consists almost exclusively of melody. It provides speakers with the ability to communicate a wide range of utterances across a distance.

The grammar of Oto-Manguean languages makes heavy use of particles. In Oto-Manguean languages, they can carry meanings associated with intonational patterns in languages like English or the Romance languages. For example, yes/no questions, topicalization, exclamation, and emphasis are indicated by particles. However, at this time, very little work has been done on Oto-Manguean sentence structure.

Oto-Manguean continues to have great import in the development of historical linguistic theory. It is a family with an internal complexity comparable to that of Indo-European but has diversified within a restricted cultural-ecological zone associated with highland agriculture. Thus, Oto-Manguean languages provide a natural testing ground for addressing the interaction of language-internal changes and regionally defined patterns of linguistic diffusion. From Archaic through Post-Classic Mesoamerica, webs of political and economic interaction mediated the divergence of related languages as villages allied themselves to larger centers emerging with the origins of the states of Pre-Columbian Mexico.

Oto-Manguean language history also involves contacts with languages outside the family but within a greater Mesoamerican culture area. Such relations include speakers of Huave, Totonac, Mixe-Zoque, Maya, and later arriving Uto-Aztecan speakers including the Aztecs. Loanwords and structural similarities shared among many of the languages of Mesoamerica support a view that networks of contact have been most important in shaping the patterns of linguistic features in this culture area. The Spanish Conquest brought a common language of the empire that must also be counted in the histories of the divergence of Oto-Manguean languages as well as in shaping our knowledge of those that were chronicled in Colonial and Post-Colonial Mesoamerica but are no longer spoken at the present time.

Today the status of Oto-Manguean languages varies from village to village. Some families have well over 100,000 speakers, but the languages of that family may range from villages of several thousand, where all the children learn an Oto-Manguean language as a first language, to villages where less than a dozen elders have any memory of their language. In most communities, bilingualism is the norm, with Spanish being learned either at home or in school. Many villages that once spoke an Oto-Manguean language have entirely shifted to Spanish. This trend is occurring in some regions at an alarming rate. Several key languages and language families are already extinct. Chiapaneco and all of Manguean have long since passed. While the documents on Manguean are sufficient to demonstrate its membership in the Oto-Manguean stock, the data were not recorded with attention to tone and other complex issues of pronunciation. Most of the Tlapanecan-Manguean languages have not been adequately documented at all. For Oto-Manguean overall, there are very few dictionaries and grammars, and few practical orthographies have been developed. The next couple of generations will decide whether Oto-Manguean voices will still be heard in Mexico's future.

References

Bartholomew, Doris A. 1960. Some revisions of proto-Otomí consonants. International Journal of American Linguistics 26. 317–29.

Campbell, Lyle, Terrence Kaufman, and Thomas C. Smith-Stark. 1986. Meso-America as a linguistic area. Language 62(3). 530–70.

de Miranda, Fernández, and María Teresa. 1995. El ProtoZapoteco, ed. by Michael J. Piper and Doris Bartholomew. Mexico D.F.: El Colegio de Mexico. 1995.

Gudschinsky, Sarah C. 1995. Lexico-statistical skewing from dialect borrowing. International Journal of American Linguistics 21. 138–49.

Hamp, Eric P. 1958. Protopopoloca Internal Relationships. International Journal of American Linguistics 24. 150–53.

Josserand, J. Kathryn, Marcus Winter, and Nicholas Hopkins (eds.) 1984. Essays in Otomanguean culture history, Nashville: Vanderbilt University Publications in Anthropology No. 31.

Kaufman, Terrence. 1994. The native languages of Mesoamerica. Atlas of the World's languages, ed. by Moseley and Asher. London, New York: Routledge.

Longacre, Robert E. 1966. On the linguistic affinities of Amuzgo. International Journal of American Linguistics 32. 46–9.

Newman, Stanley, and Robert Weitlaner, 1950. Central Otomian I: Proto-Otomi reconstructions. International Journal of American Linguistics 16. 1–19.

Rensch, Calvin R. 1976. Comparative Oto-Manguean phonology. Bloomington: Indiana University Press.

Swadesh, Morris. 1947. The phonemic structure of Proto-Zapotec. International Journal of American Linguistics 13. 220–30.

———. 1960. The Oto-Manguean hypothesis and macro Mixtecan. International Journal of American Linguistics 26. 79–111.

MARK SICOLI

See also **Mexico; Tone Languages**

P

Pacific

The area covered by the Pacific ocean occupies about a third of the Earth's surface and is by far the world's largest geographic region. It contains more than 1,000 indigenous languages, Asian mainland languages, half a dozen European languages, and a number of pidgins and creoles. Geographically, it includes the island nations and territories of Micronesia, Polynesia, and Melanesia, collectively known as Oceania. The islands occupy just less than 500,000 square miles of land area. New Guinea, the second largest island in the world after Greenland, represents 64% of this total, and New Zealand represents another 20%. The remaining 10% is divided among more than 10,000 scattered islands. Micronesia comprises the island groups east of the Philippines and north of the equator, including, for instance, the Mariana, Marshall, Caroline, and Gilbert islands. Melanesia encompasses the island groups in the southwestern Pacific extending southeastward from the Admiralty islands to Fiji, including such nations as Papua New Guinea, Solomon Islands, Vanuatu, and New Caledonia. Polynesia consists of a group of islands extending from New Zealand north to Hawaii and east to Easter Island.

Politically, the status of the islands ranges from those that are independent (e.g. Papua New Guinea, Fiji, Kiribati, Palau, Samoa—formerly Western Samoa, Tuvalu, and Tonga), those that are quasi-colonies or territories (e.g. New Caledonia, French Polynesia, and Wallis and Futuna—French overseas territories; Guam and American Samoa—United States territories; Norfolk Island—an Australian territory; Rapa Nui/Easter Island—a dependency of Chile; and Northern Marianas—a Commonwealth of the United States), those that are self-governing in free association with other countries (e.g. Cook Islands, Tokelau, and Niue in relation to New Zealand, and the Federated States of Micronesia, Chuuk, Pohnpei, Yap, and Kosrae, in relation to the United States), and those that are colonies (e.g. Pitcairn Island—a British colony), to those that have been incorporated into other nations (e.g. Hawaii—formerly a territory of the United States, now the fiftieth state).

The Pacific islands were peopled by influxes from both mainland and island Southeast Asia, with the earliest migrations to Melanesia. From Melanesia, generations of voyagers headed northward into eastern Micronesia and eastward into Polynesia. European exploration of the Pacific began with the voyages of Ferdinand Magellan in the 1520s, which inaugurated a period of Spanish and Portuguese influence, followed by Dutch, English, French, and German influences. Most of the major island groups were annexed by Britain, the United States, France, and Germany in the nineteenth century. Many Micronesian territories changed hands a number of times, beginning with Spanish colonial administration that was followed by German, Japanese, and American regimes.

The historical partition of the islands among the various European colonizers has left the islands' populations divided primarily into a Francophone and Anglophone set of nations, each with different political, economic, linguistic, and cultural orientations. More recently, however, notions of a Pacific islander identity

have begun to emerge, fueled partially by recognition of a common history of colonization, as well as by demands from nonindependent territories and nations for increased political autonomy.

The indigenous languages of the Pacific belong to two language families: Austronesian and non-Austronesian (Papuan). The Austronesian language family is the largest and most widespread family in the world. The Oceanic or easternmost subgroup of the Austronesian family is the largest and best defined of all subgroups, containing all Polynesian languages, all Micronesian languages (except Palauan and Chamorro), and all Austronesian languages of Melanesia east of the Mamberamo River in Irian Jaya (the western half of the island of New Guinea, Indonesia's easternmost province).

The Polynesian languages are a subgroup of Oceanic containing approximately 30 languages organized into two major subgroupings: Tongic (including Tongan and Niuean) and Nuclear Polynesian (all the rest). The latter is divided into two further groups: Samoic Outlier (consisting of Samoan and Tuvaluan) and Eastern Polynesian (consisting of Hawaiian, Tahitian, Maori, Rapa Nui, and Marquesan). Speakers number fewer than a million people spread across a large area of the Pacific, ranging from Hawaii in the north to New Zealand in the south and Easter Island. This so-called Polynesian Triangle encloses an area approximately twice the size of the continental United States. Despite their great geographic spread, the Polynesian languages show a relative homogeneity, which indicates that they have dispersed only in the last 2,500 years from an original center in the Tonga-Samoa area. In addition, there are some 18 further Polynesian-speaking communities known as 'Polynesian Outliers' in Micronesia (e.g. Kapingamarangi and Nukuoro, two atolls in the southwestern portion of Pohnpei state) and Melanesia (e.g. Ontong Java atoll and Rennell island in the Solomon Islands). Although some Polynesian languages have a relatively large number of speakers, such as Samoan with about 200,000, many others have very few, and some, such as Maori and Hawaiian, face possible extinction as their speakers have shifted to English.

Most of the Oceanic Micronesian languages (with the exception of Nauruan, which has not been sufficiently studied) form a closely related subgroup called Nuclear Micronesian. Palauan and Chamorro, however, appear to be the result of distinct migrations from Indonesia or the Philippines. Although Yapese is probably Oceanic, it does not appear closely related to any of the other Micronesian languages, because of its complex history of borrowing. It is also the only

Austronesian language containing glottalized consonants, (i.e. pronounced with simultaneous contraction of the vocal folds). Most of the Micronesian languages have relatively small numbers of speakers, with a few, such as Sonsorolese on Sonsoral island in Palau, having only several hundred speakers.

Historical relationships among the Papuan languages, found primarily in the New Guinea interior, an area of immense linguistic complexity, are less clear than in the much better studied Austronesian language family, and the label is best seen as a cover term for many, perhaps as many as 60, distinct families. The Papuan-speaking peoples were the earliest people in the region, occupying the Sahul continent (which later partially submerged to become the island of New Guinea) at least 40,000 years ago. By 30,000 years ago, the Bismarck Archipelago east of New Guinea was occupied by speakers of Papuan languages. Speakers of Austronesian languages, who arrived much later, established coastal communities in New Guinea, the southeastern Solomons, Vanuatu, New Caledonia, and Fiji. Contacts between speakers of Papuan and Austronesian languages have made it difficult to classify some languages as Austronesian or Papuan, such as Maisin in Papua New Guinea, which displays features typical of both language families.

The European languages spoken in the Pacific region are primarily those of the former colonizers and still have official status in most areas. Although French is important, no European language has had the impact that English has had. In addition to the standard variety of British or American English that serves as the official language in Anglophone territories, English exists in a number of locally distinct varieties in New Zealand and Hawaii, and it also exists in the form of numerous pidgins and creoles. French has left its legacy in the form of a number of local varieties of colonial French spoken, for example, in New Caledonia and Tahiti, as well as in a French creole, Tayo, in New Caledonia. Standard French still has official status in French territories of the Pacific, as well as in independent Vanuatu, formerly a condominium ruled jointly by Britain and France, where it shares co-official status with English. Because Germany lost most of its colonies after World War I, German never had a chance to spread as widely as other colonial languages. The language left some remnants, nonetheless, in the form of a German creole, Unserdeutsch, as well as borrowings, both into forms of pidgin English, such as Tok Pisin in Papua New Guinea, as well as into indigenous languages. Spanish has only a limited presence in the Pacific as the official language of Rapa Nui/Easter Island.

In addition to European languages, other languages were introduced as a result of colonization, such as Japanese, Korean, varieties of Chinese (for example, in Papua New Guinea, Nauru, Tahiti, and elsewhere), and Vietnamese in New Caledonia. Varieties of Southeast Asian languages, such as Hindi, were introduced into Fiji, and Javanese was introduced into New Caledonia. Philippine languages such as Tagalog, Ilocano, and Visayan were introduced into Hawaii through the importation of migrant labor and indentured plantation workers.

The Pacific pidgins and creoles originated primarily in the late eighteenth and early nineteenth centuries. By the latter part of the nineteenth century, trade contacts between English speakers and Pacific islanders had led to the formation of English-based pidgins spoken in various forms and with differing degrees of stability in almost the entire Pacific basin from New Guinea to Pitcairn island, from the Marshall Islands and Hawaii to New Caledonia and New Zealand. Not all of these have survived. In Micronesia and most of Polynesia (with the exception of Hawaii), forms of pidgin English disappeared because their role in internal communication was quite limited. Chinese Pidgin English, the oldest form of Pidgin English in the Pacific, which developed around the port of Canton in 1700, has subsequently died out, although a mixed form of Melanesian-Chinese Pidgin English is still used on Nauru in a variety of commercial contexts, including Chinese trade stores and restaurants, and in the phosphate mining industry, which began in 1906.

In addition, there are a number of pidgins and creoles based on indigenous languages such as Hiri Motu in Papua New Guinea, one of the three official languages of Papua New Guinea (along with English and Tok Pisin). A number of indigenous pidgins such as Pidgin Maori, Pidgin Fijian, and Pidgin Hawaiian are no longer in use.

Pidgin and creole languages based on English are more numerous than those related to any other language, attesting to the greater spread of English than any other metropolitan language. Because the French and Germans were relative latecomers to the Pacific, they found varieties of Pidgin English well established in many areas. Melanesian Pidgin English emerged on plantations in Queensland, Australia, to which workers from all over Melanesia were recruited, and later evolved into a number of distinctly named varieties, including Tok Pisin, spoken in Papua New Guinea; Bislama, spoken in Vanuatu; and Pijin, spoken in the Solomon Islands. In these highly multilingual countries, these languages are the normal everyday medium of communication for millions of people. Tok Pisin is the largest language in the South Pacific today, with as many as 2 million speakers. Most business in the House of Assembly, the country's main legislative body, is conducted in Tok Pisin, which is the most widely shared language among the members. In Hawaii, a creole English developed on plantations and is still in use, primarily among working-class speakers. There is some dispute over the status of the variety of English spoken on Pitcairn and Norfolk Islands, spoken by descendants of the British mutineers from HMS *Bounty*.

Only Tok Pisin and Bislama have received some official recognition. Tok Pisin is a de facto official language in Papua New Guinea and is spoken by more than half the population; however, English is the official medium of education. Although Bislama is recognized by the constitution of Vanuatu as the national language of the country, it is forbidden in the schools. English and French, the languages of the former colonial powers, are still the official languages of education.

The only example of a German creole is Rabaul Creole German, called Unserdeutsch, 'our German', which arose at the turn of the century as a lingua franca of the Catholic mixed-race community in Vunapope near Rabaul in what is now East New Britain Province of Papua New Guinea. This was formerly part of the German colony Kaiserwilhelmsland from 1884 to 1914. This language was apparently used by children at a mission boarding school and creolized in one generation. It is now extinct.

The only example of a French creole is Tayo, spoken in the southern part of New Caledonia around St. Louis, the site of a Catholic mission and plantation, where linguistic diversity among the surrounding tribes led to the emergence of a creolized French as a lingua franca.

References

Bellwood, Peter. 1991. The Austronesian dispersal and the origin of languages. Scientific American July. 70.

Foley, William. 1986. The Papuan languages of New Guinea. Cambridge: Cambridge University Press.

Hollyman, K.L. 1971. French in the Pacific. Current trends in linguistics, Vol. 8, Linguistics in Oceania, ed. by Thomas A. Sebeok. The Hague: Mouton.

Krupa, Viktor. 1982. The Polynesian languages: a guide. London: Routledge & Kegan Paul.

Lynch, John. 1998. Pacific languages. Honolulu: University of Hawaii Press.

Romaine, Suzanne. 1991. The Pacific. English around the World: sociolinguistic perspectives, ed. by Jenny Cheshire. Cambridge: Cambridge University Press.

SUZANNE ROMAINE

See also **Austronesian; New Guinea; Pidgins and Creoles**

Pama-Nyungan

The term Pama-Nyungan, coined by Ken Hale in the early 1960s after the words for 'man' in the northeastern and southwestern extremities of Australia, refers to a group of 160 or so languages spoken over about seven eighths of the continent (shown in gray on the map). This group was initially proposed on the basis of lexicostatistical comparisons of 100 basic words, and was believed to represent a single family of genetically related languages. The remaining eighth of the continent is considerably more diverse, and the languages were assigned to some 28 additional families, referred to collectively as non-Pama-Nyungan. It was considered that almost all Australian languages, Pama-Nyungan and non-Pama-Nyungan, are ultimately genetically related.

The lexicostatistical classification correlates well with a typological classification (the essence of which was proposed in 1940 by Arthur Capell). Most Pama-Nyungan languages are 'suffixing' (by definition, have only suffixes, and no prefixes) and dependent marking (grammatical relations signalled on noun phrases), while non-Pama-Nyungan languages are overwhelmingly 'prefixing' (by definition, have both prefixes and suffixes) and head marking (grammatical relations signaled on the verb, and sometimes also on noun phrases). There are a few exceptions: the Tangkic languages are suffixing and dependent marking but non-Pama-Nyungan; Yanyuwa is Pama-Nyungan but prefixing and head marking. The classification has been the subject of considerable controversy in recent years. There is, however, no widely accepted alternative, and most Australianists accept it in broad outline.

Current State of the Languages

Most Australian languages (Pama-Nyungan and non-Pama-Nyungan) are endangered or moribund. Many languages of the east and southeast, where conflict with Europeans was particularly intense and violent, are no longer used as vehicles of everyday communication, and are effectively dead; some disappeared before adequate records could be made of them. On the other hand, some of the most viable languages, with the best chance of future survival, are Pama-Nyungan. These include the Western Desert language, with some 4,000–5,000 speakers; Arrernte and Warlpiri, with some 3,000 speakers each; Yolngu varieties with some 2,000; and Kala Lagaw Ya (Torres Strait Islands) with some 3,000–4,000 speakers.

1 Kalibamic
2 Kukatyic?
3 Karantyic?
4 Walangama?
5 Waramungic
6 Kalkatungic
7 Yalarnngic
8 Yandic
9 Durubulic
10 Giabalic?
11 Bandjalanic
12 Gumbaynggir
13 Marawaric
14 Yotayotic
15 Yaithmathangic

Boundary of Pama-Nyungan and non-Pama-Nyungan
Group boundaries
Questionable groups
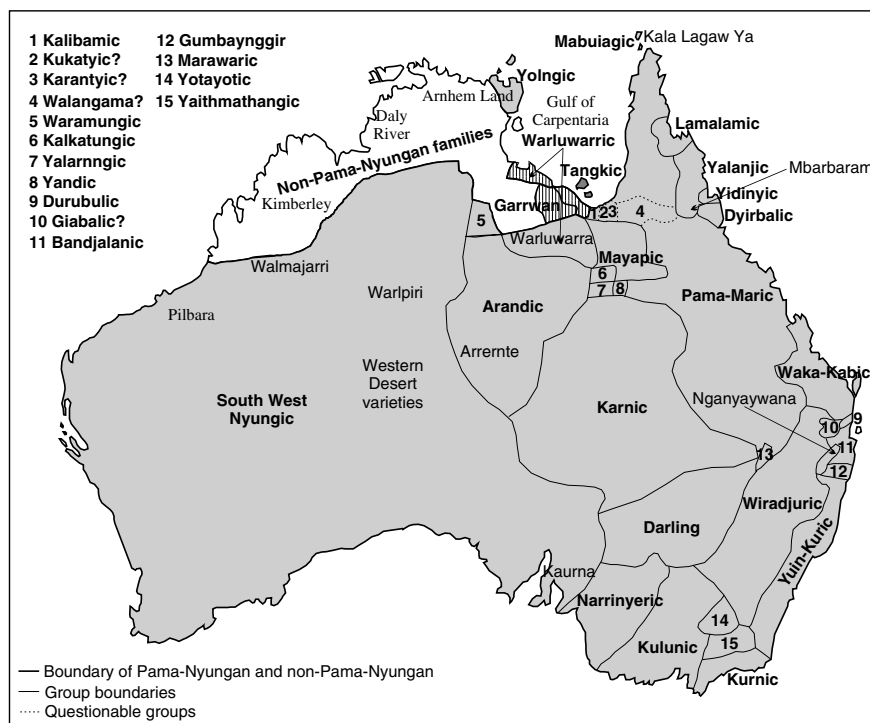
Efforts are being made by speakers and their descendants to reintroduce some dead Pama-Nyungan languages (e.g. Kaurna in the Adelaide region), and to strengthen some weakening languages (e.g. Walmajarri in the northwest). But few such efforts have achieved appreciable results, and considerably more needs to be done before the process of attrition can be considered to have been arrested generally, or for any language.

Status of Pama-Nyungan and Relation to non-Pama-Nyungan

The notion of a Pama-Nyungan family is widely accepted by Australianists; most believe that a genetic family of roughly the size shown on the map will ultimately prove valid, although we are a long way from providing convincing arguments. It has not been established as a family by rigorous application of the comparative method: little of (ancestral) proto-Pama-Nyungan phonology, lexicon, or grammar has been reconstructed, and few subgroupings have been argued on the evidence of shared innovations. One reason for this situation is that investigators have assumed the genetic relatedness of Australian languages, and have focused on the search for shared innovations, rather than attempting to reconstruct proto-Pama-Nyungan.

With improved descriptions of more languages, and use of the comparative method, the membership and scope of Pama-Nyungan have been refined somewhat since the 1960s, as have details of the subclassification. Thus, Mbabaram (Queensland rain forest region) and Nganyaywana (northeast New South Wales), originally excluded from the Australian superfamily, were shown by Robert Dixon and Terry Crowley, respectively, in the 1970s to be closely related to their neighbors; both had undergone radical sound changes that obscured cognates (related words).

Perhaps the major revision was proposed in the late 1980s by Barry Blake (1988) and Nicholas Evans (1988). They reclassified the Tangkic languages, previously considered to be Pama-Nyungan (largely on typological grounds, since these languages show low vocabulary sharing with Pama-Nyungan) as non-Pama-Nyungan, and Yanyuwa, previously taken to be a non-Pama-Nyungan isolate, as Pama-Nyungan. They assigned Yanyuwa to a subgroup with Warluwarra, spoken to the south and separated from Yanyuwa by the Garrwan languages (whose classification has also proved problematic). (These are indicated by vertical hatching on the map.) Their evidence mainly concerned forms of pronominals and case-markers.

Blake (1988) reconstructed a pronominal system for proto-Pama-Nyungan, and a different one for proto-non-Pama-Nyungan. Evans (1988) proposed further that Pama-Nyungan could be identified as a subgroup (within Australian) on the basis of specific phonetic and grammatical characteristics, for example, the wide occurrence of the variants -lu ~ -nggu as a marker for ergative case.

Dixon has always been a virulent critic of the notion that Pama-Nyungan languages constitute a genetic family of their own, lambasting the idea on methodological and other grounds. Dixon has made some valid criticisms of lexicostatistical methodology. However, the lexicostatistical classification was always regarded as a tentative means of providing a preliminary genetic classification, one that must be refined and supported by the comparative method. Most Australianists who accept the classification do so tentatively (an imperfect or faulty classification is preferable to none, provided that its provisional status is borne in mind).

Dixon also argues against Pama-Nyungan on the grounds that the putative innovations in the pronominal and case-marking systems—e.g. the first-person dual inclusive *ngali* 'we two (including you)' and the ergative variants -lu ~ -nggu—are not present in all Pama-Nyungan languages. Dixon avers that areal diffusion rather than common ancestry provides the simplest explanation for the distribution of these features. He suggests that the -nggu allomorph was independently innovated (via natural phonological conditioning) in four different places, and subsequently diffused areally. Others consider that the distribution of -nggu is better explained as a retention from a proto-language that was lost in some modern languages or groups. And as Evans (2003) pertinently observes, although instances of diffusion of specific morpheme variants are well attested in Australia, diffusion of variant sets (with complex conditioning factors) such as -lu ~ -nggu are not. In a similar way, many Australianists consider it to be probable that the absence of the first-person dual inclusive *ngali* in a few Pama-Nyungan languages represents loss of the form in these languages, rather than incomplete diffusion over the Pama-Nyungan region.

Both Barry Alpher (e.g. 1990) and Harold Koch (e.g. 2003) argue the advantages, indeed necessity, of taking inflectional paradigms into account in reconstructing proto-Pama-Nyungan, and not merely the forms of roots and affixes separately. Both show that sets of paradigmatic forms—verb inflection paradigms by Alpher (1990), and case-paradigms for singular free pronominals and interrogative particles by Koch (2003)—can be identified in a diverse selection of Pama-Nyungan languages that permit reconstruction of inflectional paradigms, and thus support the genetic unity of Pama-Nyungan.

Subgrouping

The map gives an approximate idea of the groups within Pama-Nyungan. Some 30 groups are identified, which are further divided into almost seventy subgroups. The groups and subgroups differ considerably in distribution and size. One group, the Nyungic group, takes up nearly half of the Pama-Nyungan region, and is divided into about a dozen subgroups. On the other hand, Warumingic, Kalkatungic, and Yandic consist of single languages.

The establishment of the groups and subgroups of Pama-Nyungan is in a very preliminary state, and there are doubts about the status of many. Various changes to the group and subgroup structure have been proposed since 1966. Attempts have also been made to motivate groups or subgroups by more reliable means than the original lexicostatistical criteria. In a number of cases, however, the genetic relatedness of a set of languages has been established, although not subgroup status by identification of shared innovations.

In a careful study of the Pilbara situation, Alan Dench (2001) has pointed out a variety of difficulties in determining whether shared morphological and syntactic features are innovations indicative of common inheritance or the result of diffusion. He concludes that although the languages are probably genetically related, there is no evidence that they form a subgroup of any larger group; the languages form, in his view, a linguistic area within which much lexical and grammatical diffusion has occurred. Although Dench takes the extreme position that diffusional explanations should be privileged over genetic retention, he correctly observes the difficulties inherent in distinguishing genetically inherited features from areally diffused features.

Dixon's current position (2002) is that the putative Pama-Nyungan family is made up of some 30 low-level genetic 'subgroups', which are effectively small language families on their own (the prefix 'sub-', he explains, does not imply that there is necessarily any higher order containing group, only that there may be).

Dixon further proposes that some of the accepted groups within Pama-Nyungan are areal groups, not genetic subgroups. These include the Arandic and adjacent Karnic groups of Central Australia. However, Koch (2004) and Claire Bowern (2001) provide carefully argued cases for their subgroup status on the basis of shared innovations, at least some of which (e.g. paradigmatic irregularities) are more likely to be retentions from an earlier proto-language than resulting from diffusion. Examples of such shared innovations peculiar to the Arandic languages are phonological changes (including complete loss of all initial consonants, centralization of vowels in unstressed syllables, and stress shift to the second syllable of words), and innovations in personal pronouns, nominal inflections, and interrogatives. In Karnic languages, they include lexical and morphological innovations.

Origin and Dispersal of Pama-Nyungan Languages

The uneven distribution of language families across Australia, and of groups within the Pama-Nyungan region is striking, and invites explanation. Some proposals have been put forward. These accept that proto-Pama-Nyungan was originally spoken in a much smaller homeland region, and that the continent as a whole once showed linguistic diversity comparable to that of the non-Pama-Nyungan region today. Around 3,000–5,000 years ago (the dating is a guess, based on the apparent diversity within the family), it began spreading out to ultimately cover the large region it occupies today.

Stephen Wurm (1972) suggests a link with the emergence of an advanced technology (ground axes and hafted stone tools) that also appears on the archeological horizon some 5,000 years ago. He proposes that users of this technology, speakers of proto-Pama-Nyungan, expanded out into areas formerly occupied by speakers of different languages, their advanced technology presumably giving them an edge over the local inhabitants.

A more elaborate model is suggested by Nicolas Evans and Rhys Jones (1997). These authors propose (on the basis of the diversity of subgroups of Pama-Nyungan and the number of potential sister families) a homeland region in the vicinity of the Gulf of Carpenteria. In contrast to Wurm, they do not link the spread of Pama-Nyungan to major migrations and conquests by the wielders of a new technology, but to a cluster of social changes, including the spread of ceremonial activity associated with the new highly valued technology, and the exploitation of new vegetable foods that could support large intergroup ceremonial activities. They suggest that initiation into the rituals associated with the new technology might attract payment in the form of wives for the sons of the initiators, thus giving rise to larger networks of exogamous social interactions than had existed previously. In such circumstances, proto-Pama-Nyungan could well have been a high-status language, and its usage a sign of the status of performers of the new ceremony. Thus, Pama-Nyungan could have spread out with relatively minor (though still discernible) demographic changes.

Both models are of course speculative. Nevertheless they are noteworthy in that they attempt to forge links between linguistic, anthropological, and archeological evidence. They also make potentially testable predictions concerning the distribution of bio-genetic

markers and substrate influences on modern Pama-Nyungan languages.

References

Alpher, Barry. 1990. Some proto-Pama-Nyungan paradigms: a verb in the hand is worth two in the phylum. Studies in comparative Pama-Nyungan, ed. by Geoffrey N. O'Grady and Darrell T. Tryon, 155–71. Canberra: Pacific Linguistics.

Blake, Barry. 1988. Redefining Pama-Nyungan: towards the prehistory of Australian languages. Aboriginal Linguistics 1. 1–90.

Bowern, Claire L. 2001. Karnic classification revisited. Forty years on: Ken Hale and Australian languages, ed. by Jane Simpson, David Nash, Mary Laughren, Peter Austin, and Barry Alpher, 245–61. Canberra: Pacific Linguistics.

Bowern, Claire, and Harold Koch (eds.) 2004. Australian languages: Classification and the comparative method. Amsterdam: John Benjamins.

Dench, Alan. 2001. Descent and diffusion: the complexity of the Pilbara situation. Areal diffusion and genetic inheritance: problems in comparative linguistics, ed. by Alexandra Y. Aikhenvald and Robert M.W. Dixon, 105–33. Oxford: Oxford University Press.

Dixon, Robert M.W. 2002. Australian languages: their nature and development. Cambridge: Cambridge University Press.

Evans, Nicholas. 1988. Arguments for Pama-Nyungan as a genetic subgroup, with particular reference to initial laminalization. Aboriginal Linguistics 1. 91–110.

———. 2003 Introduction: comparative non-Pama-Nyungan and Australian historical linguistics. Comparative studies in non-Pama-Nyungan, ed. by Nicholas Evans, 3–25. Canberra: Pacific Linguistics.

Evans, Nicholas and Rhys Jones. 1997. The cradle of the Pama-Nyungans: archaeological and linguistic speculations. Linguistics and archaeology: aboriginal Australia in global perspective, ed. by Patrick McConvell and Nicholas Evans, 385–417. Melbourne: Oxford University Press.

Koch, Harold. The case for Pama-Nyungan: evidence from Remains forthcoming inflectional morphology. Paper presented to 17th International Congress of Linguists, Prague, Czech Republic, July 24–29, 2003, to appear in Proceedings of the congress.

———. The Arandic subgroup of Australian languages. In Bowern and Koch, pp. 127–50.

Miceli, Luisa. Pama-Nyungan as a genetic entity. In Bowern and Koch, pp. 61–86.

O'Grady, Geoffrey N., Carl F. Voegelin, and Frances M. Voegelin. 1966. Languages of the World: Indo-Pacific fascicle 6. Anthropological Linguistics 8. 1–197.

Wurm, Stephen A. 1972. Languages of Australia and Tasmania. The Hague: Mouton.

WILLIAM B. MCGREGOR

Pāṇini

Virtually nothing is known about the life of the great Indian grammarian Pāṇini, except for the fact that he was a native of Śalātura, in northwestern India (Gandhāra); from his matronymic, we learn that his mother's name was Dākṣī. None of the numerous commentators say anything about the time in which he lived, but, based on the variety of Sanskrit he describes, he is now thought to have lived in the fourth (or maybe fifth) century BCE. His grammatical treatise, the *Aṣṭādhyāyī* ('Eight chapters'), is the most important grammatical description of Sanskrit, and a large part of the Indian grammatical tradition after Pāṇini consists of commentaries on his work.

The style of the *Aṣṭādhyāyī* appears quite striking if considered from the point of view of Western grammatical tradition: it is constituted by 3,996 *sūtra*, or aphoristic rules, by which Pāṇini not only describes grammatical rules but also defines the way in which grammatical terms and constructions are used in his metalanguage. Pāṇini's terminology constitutes his own achievement only in part: grammatical description already relied on a rather conspicuous tradition,

judging from the fact that Pāṇini mentions the names of ten of his predecessors.

The language described by Pāṇini is a variety of Sanskrit that does not correspond exactly to the language of any known text. This fact has aroused suspicion, especially among Western scholars in the nineteenth century, leading some to blame Pāṇini of describing an invented variety. A further complication is given by the fact that even the earliest commentators are at variance with Pāṇini on some points: most likely, different varieties of Sanskrit underlie their descriptions. The philological tradition of Pāṇini's text is not without problems either, as some minor parts probably belong to later commentaries (notably to Patañjali's *Mahābhāṣya* see below).

It is noteworthy that Pāṇini occasionally mentions differences between the variety he describes and earlier varieties, notably Vedic Sanskrit. This has led some scholars to state that Pāṇini had a historical view of grammar: however, in view of the existence of earlier grammarians, who had mostly described Vedic, it seems more likely that Pāṇini only wanted to mention

variants found in other treatises, without historical intentions.

Among the most important commentators of Pāṇini are Kātyāyana, who, according to tradition, lived in the third century BCE and annotated about a third of Pāṇini's work, and Patañjali, author of the *Mahābhāṣya* ('Great commentary', around 150 BCE). Patañjali's commentary, which takes as a starting point Kātyāyana's notes (which, in fact, we only know through the *Mahābhāṣya*), and adds comments, partly to the notes, and partly to Pāṇini's text, also contains in the introduction some general statements about the *Vedas* and about grammar.

Pāṇini's grammar became known to Western scholars at the end of the eighteenth century, with the beginning of Indological studies; since then, its fortune in the West has undergone quite dramatic changes. During the nineteenth century, the striking differences between the frameworks in which grammatical description was organized by Pāṇini and by the Greco-Roman tradition, together with the increasing conviction that insights into language could be gained only by means of historical investigation, progressively led to a negative evaluation of Pāṇini's work. Perhaps the worst opinions about Pāṇini were held by the American indologist Withney, who not only thought, as many of his contemporaries, that the Indian grammarian described an invented language but also dismissed his grammatical terminology as crude and philosophically not well founded.

In spite of this and other severe judgments, Indian grammatical tradition influenced Indo-European linguistics in its early development, partly providing a basis for the terminology, partly because the Indian use of listing roots helped to develop the concept of morpheme.

The twentieth century brought a complete reevaluation of Pāṇini's methods, in some cases with enthusiastic appraisals. The systematic character of Pāṇini's treatment of grammar, and the formulation of the *sūtra* as grammatical rule appealed to scholars working within various frameworks; Pāṇini was hailed as a predecessor of both structuralism and generativism. It became apparent that Pāṇini had a much more thorough understanding of phonology than his Greek contemporaries, and his description of vocalic *sandhi* was even said to presuppose a theory similar to (early) generative phonology.

Perhaps the field in which Pāṇini has enjoyed most admiration lately is case theory. In the late 1960s, after the outburst of Fillmorean Case Grammar, it was pointed out that Pāṇini's notion of *kāraka* corresponded to what Fillmore called 'deep cases', or, to state it in more modern terminology, semantic roles. Thus, there is *kāraka* of the agent, which can be expressed in the nominative or instrumental, the use of the latter ending being demonstrated by means of passive constructions.

In fact, when describing the use of 'affixes' (meaning both verbal and nominal affixes), Pāṇini I mostly has a semantically based approach, rather than an approach based on forms, as was usual within the Greco-Roman tradition. Some later grammarians gave sematically finer definitions: for example, within the philosophical school of Nyāya the *kāraka* of the agent was conceived as implying intentionality, so that only animate entities could be regarded as agents. On the contrary, Pāṇini and his followers (Pāṇinīyas) called agents all subjects of action verbs, including inanimate entities (in examples such as 'the cart goes to the village').

References

Cardona, George. 1994. Indian linguistics. History of linguistics, Vol. 1, ed. by G. Lepschy. Oxford: Oxford University Press.

Joshi, Venkateshshastri. 1980. Problems in Sanskrit grammar. Pune: 2 Dastane Ramchandra & Co.

Rocher, Rosane. 1975. India. Current trends in linguistics, Vol. 13, ed. by T. Sebeok. The Hague, Paris: Mouton.

Vasu, Śrīśa a Chandra. 1962. The Aṣṭādhyāyī of Pāṇini Allahabad 1891, repr. Delhi: Motilal Banarsidass.

SILVIA LURAGHI

Papiamento

Papiamento (Papiamentu) is a language spoken north of Venezuela on the Leeward Islands of the Netherlands, Antilles, Curaçao, and Bonaire, and on the island of Aruba, otherwise known as the ABC islands. It is estimated that Papiamento is spoken natively by about a quarter of a million people, mostly in the Caribbean, and by some 40,000 people living in the Netherlands. Although the official language on the ABC islands is Dutch, Papiamento pervades all levels of society, including the educational system and

government. Papiamento also has a rich literary tradition, an established presence in electronic and digital media, and several newspapers and magazines are also published in the language. The name is derived from the verb *papia* meaning 'to speak' and the suffix, -mento. Thus, Papiamento can be literally interpreted as '(the act of) speaking'.

Papiamento is a creole language, which emerged during the second half of the seventeenth century around the period of European colonization of the Caribbean islands. Like other creole languages, Papiamento inherited structural characteristics from a mixture of West-African languages (the substrate languages). Approximately two thirds of its vocabulary comes from Iberian languages (the superstrate languages), about a quarter from Dutch, and the rest from other languages including French and English. The language also has traces of Amerindian and West African vocabulary as well.

The history of the ABC islands provides some insight into the social and linguistic context in which Papiamento developed: originally populated by Amerindians speaking an Arawakan language, Curaçao was settled by the Spanish in 1527. After a century of Spanish occupation, the Dutch conquered Curaçao in 1634 at which time most of the Amerindians and the Spanish were driven from the island. Shortly thereafter, it became a depot for the slave-trading practices of the Dutch West-India Company. Not long after the fall of the Dutch Brazilian Empire in 1654, a substantial population of Portuguese Jews (and non-Jewish settlers as well) arrived on the island. Despite the diverse linguistic backgrounds of the island's inhabitants, Papiamentu is believed to have stabilized as a creole on Curaçao around 1700 and spread to Bonaire and Aruba by the end of the century.

Although this brief account of the islands' history demonstrates the importance of Dutch and Portuguese in the incipience of the language, Papiamento clearly exhibits Spanish elements and in fact, a diary belonging to a Jesuit missionary on Curaçao dating from 1704, first referred to the language on the island as 'broken Spanish'. This has caused some controversy surrounding the origins of the language with respect to whether it is actually a Portuguese- or Spanish-based creole. Some scholars argue that the Portuguese contribution to the language has been obscured by its continual contact with Spanish. Consequently, the earliest forms of Papiamento are thought to have exhibited the Portuguese contribution more so than the Spanish; however, the earliest written attestation of the language, a letter dating from 1776 written by a Portuguese Jew to his wife, shows a remarkable resemblance to the language spoken on the islands today, replete with elements of the Spanish language thought to be less prominent in the earlier stages of Papiamento. Thus, questions concerning the precise origins of Papiamento continue and as it is not directly related to any other Portuguese- or Spanish-based creole, scholars can only continue to piece together linguistic and historical evidence to help shed light on the genesis and subsequent evolution of the language.

Papiamento has existed in tandem with Dutch on the islands for centuries; however, the Dutch influence in the language is far less obvious than the Iberian influences, in both historical and contemporary contexts. In spite of the fact that Dutch is the official language of the regions where Papiamento is spoken natively, it is quickly becoming far less common even in the areas of government administration, education, and the judiciary system, realms of society in which it was traditionally used, testifying to the high level of social prestige affiliated with Papiamento—a fact that makes it somewhat unique among creole languages. Papiamento's social currency is also demonstrated by the fact that it has an official orthographic system, differing slightly on the islands of Curaçao and Bonaire than on the island of Aruba. These spelling differences can be seen in the name of the language itself, while the former islands prefer Papiamentu as opposed to Papiamento, its Aruban counterpart. Additionally, after careful language planning efforts, in 1987, Papiamento was officially integrated into the educational system of the ABC islands.

The phonemes (sounds) of Papiamento include almost all the consonants and vowels found in American Spanish, as well as some Dutch vowels. A linguistic trait of Papiamento not found in either the Iberian languages or Dutch, and which is rare in creole languages in general, is the use of high (´) and low (`) tones to distinguish between pairs of words (eg. *párà* 'bird', *pàrá* 'to stand, to stop'; *wárdà* 'guard service', *wàrdá* 'to wait'). This feature of the language is attributed to the West African, or substratal influences. West African structural influences typical of creole languages can be noticed in serial-verb constructions such as *Ela **kore bai** su kas* or 'She or He ran (go) home,' or focus particles found in sentence-initial positions like ***Ta** e buki m'a dunabu* or '(Focus) I gave you the book.' Below are some Papiamento words that illustrate the Iberian and Dutch contributions to the language:

Portuguese: *bai* 'to go'; *preto* 'black'; *bringa* 'to fight' (derived from brigar)

Spanish: *Dios* 'God'; *salida* 'departure'; *ruman* 'brother' (derived from hermano)

Dutch: *wikent* 'weekend'; *spiel* 'mirror' (derived from spiegel), and *wak* 'to see' (derived from waken).

References

Agard, Frederick B. 1985. Papiamentu grammars and the structure of the language. For Gordon H. Fairbanks, ed. by Veneeta Z. Ascon and Richard L. Leed. Honolulu: University of Hawaii Press.

Anderson, Roger W. 1990. Papiamentu tense-aspect, with special attention to discourse. Pidgin and Creole Tense-Mood-Aspect systems, ed. by Jon Singler. Amsterdam: John Benjamins.

Arends, Jacques, Pieter Muysken, and Norval Smith (eds.) 1994. Pidgins and Creoles: an introduction. Amsterdam: John Benjamins.

Goilo, E.R. 1997. Papiamentu textbook. Aruba: VAD Printers.

Kouwenberg, Sylvia, and Eric Murray. 1994. Papiamentu. Newcastle: Lincom Europa.

Martinus, Frank. 1990. Papiamentu: the road to emancipation. Language reform: history and future, ed. by István Fodor and Claude Hagège. Hamburg: Helmut Buske Verlag.

Wood, Richard E. 1972. New light on the origins of Papiamentu: an eighteenth-century letter. Neophilologus 56.

BETSY BARRY

See also **Caribbean; Pidgins and Creoles**

Paralanguage

The term paralanguage refers to a wide variety of nonverbal behaviors relevant to communication. Vocal paralanguage includes voice quality, intonation, and voice dynamics such as loudness, pitch range, and rate of speech. Bodily paralanguage incorporates facial expression, hand gestures, posture, touch, and physical proximity to others. These behaviors, all of which may co-occur with verbal communication, differ from language per se in that they do not form part of a hierarchically structured, coded system in which a discrete symbol is arbitrarily associated with a particular meaning.

While paralanguage does not convey meaning in the same way as language, it plays important functions in communication. Paralinguistic behavior conveys the emotional state of speakers and listeners, expresses semantic nuances that could not be inferred from language alone, helps regulate the flow of interaction, and provides cues regarding the sex, age, culture, and socioeconomic status of speakers.

The study of paralanguage has grown significantly over the past 30 years. Technological advances such as audio and videorecording, and more recently, the widespread availability of computers and specialized software analysis packages, have facilitated a more systematic approach to the study of paralanguage. The sections below describe some key paralinguistic channels: voice quality, intonation, and hand gestures (see Siegman and Feldstein (1987) and Feldman and Rime (1991) for an overview of these and other paralinguistic channels).

Voice Quality

Voice quality may be defined as a quasipermanent setting of the vocal tract that characterizes the speech of an individual, or a cultural, social, or linguistic group (see Laver (1980) for a thorough introduction to this subject; and Kent and Ball (2000) for an overview of recent research).

Voice quality settings may be roughly divided into two categories: laryngeal and supralaryngeal. Laryngeal settings (also known as 'phonation types') involve different patterns of vocal fold vibration. These settings include modal voice, breathy voice, creaky voice, harsh voice, falsetto voice, whisper, and the various possible combinations of these phonation types (e.g. harsh whispery voice). Modal voice involves quasiperiodic vibration of the vocal folds, in that the vocal folds are open and closed for approximately equal periods of time, lending the voice a smooth, even quality. Other laryngeal configurations may be thought of as departures from this 'neutral' setting. For example, for a breathy voice, the vocal folds are open longer than they are closed, producing audible friction. By contrast, for a harsh voice, the vocal folds are closed longer than they are open, and the vibration is irregular, producing a rasping, uneven quality that is often enhanced by the vibration of the ventricular or 'false folds'. For a whisper, the vocal folds do not vibrate at all: the vocal folds are closed along most of their length, and a triangular opening is left at the posterior part of the vocal folds, generating strong friction as air passes through. Some laryngeal settings are readily identifiable by their pitch, as, for instance, the very high pitch of falsetto voice, involving the vibration of a small portion of the vocal folds, and the very low pitch of creaky voice, in which the listener can hear each separate vibration of the vocal folds.

Supralaryngeal settings involve different configurations of the larynx, pharynx, soft palate, tongue, lips, and jaw, all of which affect the resonance of the voice.

The larynx may be raised or lowered, the pharynx widened or constricted, the soft palate raised or lowered, the tongue and lips protruded or retracted, and the jaw lowered or raised—all of which, separately or in combination, have distinct acoustic and auditory effects. For example, a speaker with a habitually lowered soft palate will have a nasal voice. A speaker with a lowered larynx will have a deeper sounding voice than he or she would with a neutral or raised larynx, by virtue of having produced a longer vocal tract with correspondingly lower formant frequencies.

The combination of various laryngeal and supralaryngeal settings contributes to the characteristic 'tone of voice' or 'accent' that listeners come to associate with different individuals and regional or social groups. For example, North American females are more likely than their male counterparts to speak in a habitually breathy voice—a habit that some women may deliberately alter in order to sound more professional in a particular work or social setting. Certain regional accents of English may be distinguished by a particular voice quality setting, such as, for instance, the nasal quality characteristic of speakers in the southern United States of America. Some foreign accents are stereotypically associated with a particular voice quality setting, such as the quasi-permanent retraction of the tongue tip (or 'retroflexion') that often characterizes an East Indian accent.

In addition to suggesting social, cultural, and linguistic characteristics of speakers, voice quality settings can be deliberately exploited by a single individual in different communicative settings to convey paralinguistic meaning—whether to signal the speaker's actual emotional state or attitude toward what he or she is saying. For example, in English-speaking countries, a whisper is often used to signal secrecy, while a breathy voice may be used to convey intimacy or sexual attractiveness. A creaky voice may sometimes be used to suggest sadness or boredom, while lowered larynx voice may be used by speakers who wish to sound serious or authoritative.

Intonation

Intonation refers to the variations in pitch or fundamental frequency that occur in speech, creating its characteristic melodies. In intonational analysis, these melodies are broken down into 'tone units' or 'intonation groups'. Tone units usually consist of five or six words, often constituting a grammatical constituent such as a sentence, clause, or phrase, but sometimes consisting of a single word, or even a single syllable. The major characteristic of a tone unit, independent of its length or grammatical status, is that it contains one major pitch movement, realized primarily on the most prominent, or 'accented' syllable in the tone group. Within the British system of intonation analysis, there are five major nuclear tones: falls, rises, levels, fall–rises, and rise–falls. Within the predominantly American autosegmental metrical system, (for example, the ToBI, or 'Tone and Break Index' system), these and other contours are expressed as different patterns of high and low tones ('H' and 'L'). (See Cruttenden (1997) for a thorough overview of intonation in the British and American traditions.)

Intonation may have linguistic or paralinguistic meaning. For instance, in English, an utterance produced with falling intonation is associated with a statement or declarative sentence, while a rising intonation is often used for questions. Such uses of intonation are often considered linguistic. However, there are many intonational variations that carry paralinguistic meaning. For instance, some speakers use rising intonation for declarative sentences as a way of checking whether or not the listener understands what they are saying, to suggest openness to an alternative point of view, or, possibly, to convey a lack of confidence in what they are saying. This particular usage is more common among young North American females than their male counterparts. The word 'yes', spoken slowly with a rising then falling pitch, may suggest doubt or hesitation on the speaker's part, even though the word itself conveys agreement. The same word, spoken slightly more quickly, with a falling then rising pitch, may suggest impatience or irritation on the part of the speaker. These are but a few of many possible examples of the paralinguistic use of intonation.

Intonation, in combination with voice quality, and other features of voice dynamics, such as variations in pitch range (wide vs. narrow), intensity (loud vs. soft), and rate of speech (fast vs. slow), provide the listener with significant cues regarding the speaker's attitude and/or emotional state. For example, a wide pitch range is often characteristic of a happy mood, and a more restricted pitch range is characteristic of a more depressed state. Anger is often associated in a loud, high-pitched voice, while fear may be expressed with a quieter, high-pitched voice, sometimes combined with a whispery quality. In a state of excitement, people often speak more quickly than when they are calm. The qualities associated with a particular mood may be characteristic of a speaker's habitual use of the voice, possibly suggesting more permanent personality traits.

Gestures

While speaking, people can often be seen moving their hands up and down, pointing, or drawing pictures in the air, in ways that appear connected to what they are saying. These movements, as opposed to other hand

movements that they might happen to make while speaking, such as smoothing their hair or removing lint from their pants (the latter are called "self-adaptors"), are considered communicative gestures.

Adam Kendon—a pioneer in the research on gestures—has characterized hand gestures as existing along a continuum, ranging from completely nonlinguistic gestures to the fully linguistic hand movements that comprise actual sign languages (see McNeill (1992:37) for a fuller description of Kendon's continuum). In the middle of this continuum are the semilinguistic hand gestures known as 'emblems'. Emblems have a fixed form that is associated with a particular meaning, and may in this sense, be considered to be quasilexical. Emblems vary considerably from culture to culture. For example, the 'thumbs-up' gesture often used in English-speaking countries to signal a victory or a generally positive attitude toward a particular situation, has an obscene connotation in Iran. Some cultures—Neopolitan Italian culture, for example—have elaborate sets of emblems that are extensively used in day-to-day conversation. Because of their quasilexical nature, emblematic gestures can often be used instead of words in face-to-face interaction for communicative effect.

However, the majority of communicative hand gestures are nonlinguistic, and co-occur with speech. Nonlinguistic gestures have been categorized in a variety of ways over the past 50 years; but the categories developed by McNeill (1992) are those most commonly used in the current literature on gestures. Nonlinguistic gestures may be classified as iconic, metaphoric, deictic, or beats. Iconic gestures illustrate concrete entities in the world, while metaphoric gestures depict abstract concepts. Iconic and metaphoric gestures have no fixed form, and are spontaneously invented by speakers to physically represent some aspect of what they are saying. For example, while saying 'and the guy just grabbed her purse', a speaker might physically depict this action in an iconic gesture—thereby often conveying information not fully expressed in words, such as the precise movement involved in the grabbing. Deictic gestures involve pointing, either to a location within the speaker's environment, or to an area in space that symbolizes a location referred to in the accompanying speech. Beats are short, rhythmic movements of the hands, often physically oriented toward the listener (e.g. involving an open palm facing the listener) and do not represent anything in the accompanying speech. Rather, they are used to emphasize the delivery of information, help structure discourse, and help regulate aspects of interaction, for instance, turn-taking routines.

As noted above, nonlinguistic gestures are closely synchronized with speech. Gestures may be divided into three phases: a preparatory or onset phase, in which the hands rise from a position of rest; a stroke phase, the most 'energetic' portion of the gesture, or the point at which the gesture appears to have reached its 'destination'; and a recovery or offset phase, the point at which the hands return to a position of rest. The stroke of a gesture almost always co-occurs with the word or phrase with which it is meaningfully connected. Moreover, the stroke often co-occurs with the most prominent, or accented syllable in a tone unit, suggesting that the planning of gesture is closely linked with the planning of speech.

In recent years, there has been a surge of interest in the study of gesture and other paralinguistic phenomena, such as facial expression, voice quality, and intonation. Traditionally, these features have often been considered to reflect the emotional state of speakers. While this approach continues to form a significant part of research on paralanguage, recent research has also highlighted the important semantic information conveyed through paralinguistic channels (see, for example, McNeill 1992, 2000). Moreover, there is a growing interest in the complex ways in which linguistic and paralinguistic channels interact. This line of research is leading to an enriched understanding of communication as integrated, multimodal behavior— and perhaps challenging the traditional distinction between language and paralanguage.

References

Armstrong, David, William Stokoe, and Sherman Wilcox. 1995. Gesture and the nature of language. Cambridge, New York, and Melbourne: Cambridge University Press.

Birdwhistell, Ray. 1970. Kinesics and context. Philadelphia: University of Pennsylvania Press.

Cruttenden, Alan. 1986. Intonation. Cambridge, New York, and Melbourne: Cambridge University Press. 2nd edition, 1997.

Feldman, Robert, and Bernard Rime (eds.) 1991. Fundamentals of nonverbal behavior. Cambridge, New York, Port Chester, Melbourne, and Sydney: Cambridge University Press, and Paris: Editions de la Maison des Sciences de l'Homme.

Kent, Raymond, and Martin Ball (eds.) 2000. Voice quality measurement. San Diego, California: Singular Publishing Group.

Laver, John. 1980. The phonetic description of voice quality. Cambridge: Cambridge University Press.

McNeill, David. 1992. Hand and mind. Chicago, and London: University of Chicago Press.

———. (ed.) 2000. Language and gesture. Cambridge, New York, Melbourne, and Madrid: Cambridge University Press.

Poyatos, Fernando. 1993. Paralanguage. Amsterdam, and Philadelphia: John Benjamins Publishing Company.

Siegman, Aron, and Stanley Feldstein (eds.) 1987. Nonverbal behavior and communication. Hillsdale: Lawrence Erlbaum Associates.

ALLISON BENNER

See also **Communication Theory; Conversation Analysis; Emotion and Language; Signed Languages**

Parsing

A recognizer is an algorithm that takes a string as input and either accepts or rejects it, depending on whether or not the string is a sentence of a particular grammar (Grosz et al. 1986). A parser is a recognizer that also shows all the ways in which the string can be derived, in the form of the invoked grammar rules arranged in trees or nested square brackets.

The most common way to represent a grammar is as a set of production rules, which are also called rewrite rules. Phrase structure grammars are sets of production rules that specify the various ways in which a sentence can be decomposed into its constituent syntactic units, which themselves can be decomposed into smaller units, and finally into individual words. The production rule S → VP NP shows that a sentence can consist of a noun phrase followed by a verb phrase, and the rule ADJ → small|red|new shows that an adjective can consist of one of the following individual words: small, red, or new.

Symbols that are themselves further expanded by rules are called nonterminal symbols, which generally correspond to syntactic classes. Symbols that correspond to individual word strings that must be found in the input sentence are called terminal symbols (Rich and Knight 1991). The nonterminal symbol S, which stands for the input sentence in its entirety, is called the root of the grammar. An example of a simple grammar which derives the sentence 'John ate the cat' is given by Allen (1995:42).

(1) S → NP VP
(2) VP → V NP
(3) NP → NAME
(4) NP → ART N
(5) NAME → John
(6) V → ate
(7) ART → the
(8) N → cat

The parse tree corresponding to this parse is shown in Figure 1.
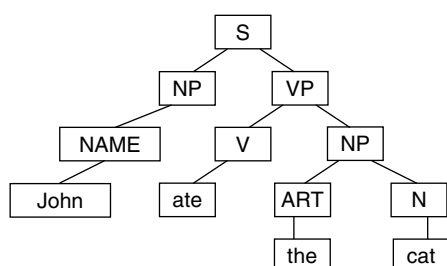
Figure 1. A tree representation of John *ate the cat*.

Context-Free Parsers

The term context free (CF) was introduced by Chomsky in 1956. CF grammars consist of rules with a single symbol on the left-hand side. Such rules are always applicable, since they are in no way dependent on the nature of constituents elsewhere in the input text. Most of the CF parsing algorithms were developed in the 1960s for use with computer language compilers, but pure CF grammars are not effective for describing natural languages. Grammars that are not CF are said to be context sensitive, where the rules have been augmented with additional conditions, meaning that they only apply if certain things are found elsewhere in the input text.

Top-Down and Bottom-Up Parsing

A top-down parsing strategy starts with the start symbol S and then searches through different ways to rewrite the symbols until either the input sentence is reproduced or all different ways to rewrite the symbols have been tried. A top-down parse using the simple grammar above to derive the sentence 'John ate the cat' would proceed as follows:

S

➜ NP VP (rewriting S)
➜ NAME VP (rewriting NP)
➜ John VP (rewriting NAME)
➜ John V NP (rewriting VP)
➜ John ate NP (rewriting V)
➜ John ate ART N (rewriting NP)
➜ John ate the N (rewriting ART)
➜ John ate the cat (rewriting N).

The bottom-up parsing strategy, on the other hand, starts with the actual words of the sentence and uses the rewrite rules in reverse until either the S symbol alone is generated or all combinations of the rewrite rules have been tried. A bottom-up parse of 'John ate the cat' might proceed as follows:

John ate the cat

➜ NAME ate the cat (rewriting John)
➜ NAME V the cat (rewriting ate)
➜ NAME V ART cat (rewriting the)
➜ NAME V ART N (rewriting cat)
➜ NP V ART N (rewriting NAME)
➜ NP V NP (rewriting ART N)
➜ NP VP (rewriting V NP)
➜ S (rewriting NP VP)

LR parsers read their input from left to right and produce a rightmost derivation, i.e. expand the rightmost non-terminals of the production rules first.

Chart Parsers

In a standard parser, if an early segment can be analyzed in more than one way, each step in the recognition of subsequent phrases in the sentence will be repeated once for each earlier analysis. Chart parsers (Winograd 1983) provide a means of eliminating such redundant computation. Kay described the chart as 'a kind of well formed substring table'. It allows the parser to store the partial results of the matching it has done so far so that this work need not be duplicated.

First, we will consider the bottom-up chart parser. Each word in the input sentence is considered in turn. The current input word is called the key. For each new key, we look for rules that match a word sequence involving the key—either a rule that begins with the key, or a rule that has already been started by earlier keys, which can either be extended or completed by the current key. Allen presents the following example of a grammar:

(1) S → NP VP
(2) NP → ART ADJ N
(3) NP → ART N
(4) NP → ADJ N
(5) VP → AUX VP
(6) VP → V NP

If we are parsing a sentence that begins with ART, ART becomes the first key, and rules (2) and (3) are matched since their right-hand sides both begin with ART. The following notation is used to show that rules (2) and (3) can be continued from the point of the ART (denoted by 'o') onwards. The partially matched rules are stored in the chart, along with their locus (starting position in the sentence) and position of the last word to match so far:

(2′) NP → ART o ADJ N (1,2)
(3′) NP → ART o N (1,2)

If the next key is ADJ, then rule 4 can be started, and rule 2′ can be extended:

(2′) NP → ART ADJ o N (1,3)
(4′) NP → ADJ o N (2,3)

The chart keeps the record of all the rules found so far which have been matched partially by the previous keys. These partially matched rules are called active arcs.

Active arcs are never removed from the chart; hence, their original and extended forms are simultaneously present. The chart also records the completely matched syntactic constituents (as defined by the grammar rules) found so far in the parse.

Completed constituents are placed on a list called the agenda. The algorithm proceeds by selecting a completed constituent from the agenda. For each matching rule in the grammar, an active arc is combined with the completed constituent in the agenda. Whenever the agenda is empty, all possible syntactic interpretations of the next word in the sentence are added. The parse ends when an active arc and a completed constituent combine to produce an entire sentence S. If all possible parses of a sentence are required, the process continues until the agenda is empty. The top-down chart parser is called the Earley parser (Earley, 1970).

Augmented Transition Networks

An early model for natural language grammars was a transition graph called the finite state machine. This consists of a network of nodes (states) and directed arcs or transitions, marked by arrowheads pointing away from the previous state and toward the next state. Each arc has a symbol, representing terminal symbols (individual words such as 'red' or syntactic classes such as 'adjective') which may appear in an input sentence. There may be more than one arc arriving at each state or leaving each state.

There is one distinguished state called the start state (with no inward arcs), and a number of distinguished states called final states (with no outward arcs). Starting at the start state, the words of the sentence to be parsed are read off in turn. If the first word read off matches the symbol on one of the arcs leaving the start state, the next state is the one found by following that arc. The second word is read in, and if that word matches an arc leaving the second state, the arc is followed to the third state. This process continues, and if we have arrived at one of the final states as a result of reading in the last word, the sentence has been accepted by the augmented transition network (ATN) grammar. Sometimes an arc can return to the very same state that it left, to show that optional additional words of the type specified by the arc's symbol are permitted at that point in the sentence.

ATN are finite state machines augmented by allowing not only terminal symbols on the arcs, but nonterminal symbols representing entire phrases (Woods 1970). These phrases are themselves represented by ATNs. For example, an ATN designed to recognize a prepositional phrase is shown in Figure 2.

Figure 2. A simple augmented transition network.

To make the transition from S2 to finish, there must be a noun phrase (NP) in the input sentence. In an operation called a 'push', processing is suspended at the point the NP is requested, and a new ATN is entered, which accepts subsequent words from the input sentence. This is analogous to a subroutine call in a computer program. When the final state of the new ATN is reached (showing that the sequence of input words since entering the new ATN was indeed an NP), an operation called a 'pop' takes place and processing of the input sentence resumes at the point in the original ATN from where it left off. A called ATN may itself call other ATNs, and an ATN may call itself (recursion). The record of the sequence in which the ATNs have called each other is called a 'stack', analogous to a stack of plates, each plate representing an ATN, which may be 'pushed' onto or 'popped' off the stack. The most recently called ATN is at the top of the stack, but it cannot be 'popped' off until any ATNs it calls itself are completed. Advantages of ATNs are their generative power, efficiency of representation, and their ability to capture regularities in language.

Probabilistic Parsing

A deterministic parser consists of a set of rules, each with equal a priori likelihood. It either generates just one parse according to the ordering of the rules, or all possible parses of a sentence, but does not mention whether any one of these parses is any likelier than the others. A probabilistic parser, on the other hand, may offer alternative rules for the decomposition of the constituents of a sentence, each with an associated probability value, e.g. VP → V NP 0.6, VP → V PP 0.4. The overall probability of a parse is found by multiplying together the probabilities of all the rules that were called upon in the derivation of that parse. The likeliest parse among many possible parses is found by the Viterbi algorithm. The strategy of exploring the paths with high probability constituents first is called Best First Parsing.

The Penn Treebank contains over two million words of American English parsed manually. It has proved useful as a reference corpus for the comparison of different parsing algorithms, such as in the PARSE-VAL evaluations (Black et al. 1991). In 1994, Magerman was able to derive rules automatically for a probabilistic parser through an analysis of the Penn Treebank. The simplest approach is to count the number of times each rule is used in the treebank and normalize this to estimate the probability of each rule.

For certain grammars, the most probable parse cannot be found easily using the Viterbi algorithm. In such cases, a most probable parse may be found using Monte Carlo simulation techniques. Many parse trees for the same sentence are randomly generated. The likelihood of each rule being used in a particular random parse depends on its probability. When the sample is large enough, the most frequently occurring output parse tree over all the random tests is assumed to be the most probable parse.

Tagging

Tagging is marking items in a text with additional information, related to their linguistic properties. This information is attached to the text items in the form of codes called tags. Here, we will mainly discuss the automatic assignment of part of speech (POS) category tags to words in a text; but tagging can also involve the assignment of semantic categories as a by-product of word sense disambiguation.

The Brill (1995) tagger operates as follows. An initial guess is made as to the POS of each word in the text by an initial state annotator. Simple methods of doing this include (a) labeling all words with their most frequent tag as found in the training corpus and (b) assuming all unknown words are nouns. The output of the initial state annotator is then compared with the manually annotated Penn Treebank, and a list of transformations is learned that can be applied to the output of the initial state annotator, according to the nature of words in the vicinity of each tagged word. The transformations have two components, a rewrite rule (such as 'change the tag from modal to singular noun') and a triggering environment (such as 'if the preceding word is a determiner'). Using this transformation, if the word 'can' has been given an incorrect initial tag as in 'the_DT can_MD rusted_VBD', the modal (MD) tag is replaced by a singular noun (NN) tag, because the previous word has a determiner (DT) tag. Other transformations are 'change an infinitive verb to a singular noun if one of the previous two tags is a determiner', and 'change a singular noun to a verb infinitive if the previous word is 'to'.

The CLAWS probabilistic tagger (Garside et al. 1987) takes an input sentence such as 'Henry likes stews'. and compares it word by word against the CLAWS lexicon, which shows that 'Henry' must be tagged as a proper noun (NP); 'likes' and 'stews' could either be tagged as plural nouns or verbs (NNS or VBZ), and the full stop always has its own tag (.). For each of the four possible tag sequences spanning this region of ambiguity, a value is generated by calculating the product of the frequencies per thousand for successive tag adjacencies as found in the Brown corpus, as shown below. The number 17 shows that in the Brown corpus, the tag NP is followed by the tag NNS 17 times per 1,000, while 983 times out of 1,000 it is followed by something else.

Value(NP-NNS-NNS-.) = $17 \times 5 \times 135 = 11{,}475$
Value(NP-NNS-VBZ-.) = $17 \times 1 \times 37 = 629$
Value(NP-VBZ-NNS-.) = $7 \times 28 \times 135 = 26{,}460$
Value(NP-VBZ-VBZ-.) = $7 \times 0 \times 37 = 0$

Thus, the tag sequence 'Henry_NP likes_VBZ stews_NNS ._' is found to be the likeliest. Since there are generally too many possible sequences to find a value for them all, the Viterbi algorithm (Forney 1973) is used, which means that only the most promising sequences are tested. The CLAWS tagger relies on bigram (pairs of adjacent tags) frequencies, while in 1988, Church described a tagger based on trigram (sequences of three adjacent tags) frequencies.

Affix Stripping

Affix stripping and morphological parsing are parsing at the subword level. Instead of analyzing a sentence in terms of its constituent words, a word is analyzed in terms of its constituent morphemes. Klavans and Chodorow (1991) used an instructional morphological parser to teach morphological theory. Several sets of rules exist for the automatic removal and replacement of common suffixes (stemming), including those produced by Paice (1990), Lovins, and Porter. Recognizing suffixes automatically can help in POS identification, where, for example, any word ending in 'ation' must be a noun. The simplest form of stemming is the reduction of all regular nouns and verbs to their inflectional root, which is the singular form for nouns and the infinitive form for verbs.

The rule sets for the removal and replacement of common suffixes are in the form of production rules, in the form 'old suffix → new suffix', meaning that if a word ends in the old suffix, this must be removed and replaced by the new suffix. The new suffix may be null, meaning that the old suffix is removed and is not replaced with anything. The rules must be presented in a fixed order. Four of the rules given by Paice are as follows:

(1) ING → NULL
(2) E → NULL
(3) ION → NULL
(4) OLV → OLUT

Imagine that the word 'resolving' is given to each rule in turn. Rule (1) removes the suffix 'ing' and replaces it with nothing; hence, we are left with 'resolv'. This does not end in 'e' or 'ion', hence, rules (2) and (3) have no effect. However, rule (4) states that the current suffix 'olv' must be removed and replaced by 'olut', leaving the string 'resolut'. The words 'resolve' and 'resolution' will also be reduced to 'resolut', rendering all three words equivalent. Prefix removal may be less useful than suffix removal, as the removal of a prefix may radically alter the meaning of a word: consider the group 'bisect', 'dissect', 'insect', 'sect'.

Semantic Parsing

A semantic grammar, like a syntactic grammar, consists of production rules, but instead of specifying only the syntactic constituents that may be present in a sentence, may also specify domain-specific semantic categories. Allen (1995) gives an example in the domain of airline bookings, of a semantic grammar that should cater for such input phrases as 'flight 457 to Chicago'. This would include such rules as 'FLIGHT-NP → FLIGHT-N NUMB' (A noun phrase referring to a flight consists of a noun referring to a flight and a number) and 'FLIGHT-NP → FLIGHT-N to CITY-NP' (A noun phrase referring to a flight consists of a noun meaning 'flight', followed by 'to', then a noun phrase which is the name of a city). Different algorithms for semantic parsing vary according to whether it is necessary to produce a syntactic parse first, or whether one should proceed directly with the semantic parse.

References

Allen, James. 1995. Natural language understanding. Redwood City, CA: Benjamin Cummings.

Black, Ezra et al. 1991. A procedure for quantitatively comparing the syntactic coverage of English grammars. Proceedings of the fourth DARPA Workshop on Speech and Natural Language. San Mateo, CA: Morgan Kaufman.

Brill, Eric. 1995. Transformation-based error driven learning and natural language processing: a case study in part-of-speech tagging. Computational Linguistics 21(4). 543–65.

Earley, Jay. 1970. An efficient context-free parsing algorithm. Communications of the ACM 13(2). 94–102.

Forney, G D Jr. 1973. The Viterbi Algorithm. Proceedings of the IEEE 61(3). 268.

Garside, Roger, Geoffrey Leech, and Geoffrey Sampson (eds.) 1987. The computational analysis of English, a corpus-based approach. London: Longman.

Grosz, Barbara J, Karen Spärck Jones, and Bonnie Lynn Webber (eds.) 1986. Readings in natural language processing. Los Altos, CA: Morgan Kaufmann.

Klavans, Judith, and Martin S Chodorow. 1991. Using a morphological analyser to teach theoretical morphology. Computers and the Humanities 5. 281–87.

Paice, Chris D. 1990. Method for evaluating stemming algorithms based on error counting. Journal of the American Society for Information Science (JASIS) 47(8). 632–49.

Rich, Elaine, and Kevin Knight. 1991. Artificial intelligence. New York: McGraw-Hill.

Winograd, Terry. 1983. Language as a cognitive process: syntax. Reading, MA: Addison-Wesley.

Woods, W A. 1970. Transition network grammars for nnatural language analysis. Communications of the ACM 13(10): 591–606.

MICHAEL P. OAKES

See also **Phrase Structure**

Paul, Hermann

Hermann Paul was one of the most influential linguists of the late nineteenth and early twentieth centuries. Paul was a neogrammarian (Germ. *Junggrammatiker*), a school of linguistics that flourished in his time and had an immense influence on linguistics thereafter, as discussed further below. His interest in languages and linguistics dates back to his schooldays, when he developed a taste for medieval German literature and Middle High German (superseding an earlier interest in mathematics).

In a brief autobiographical essay published just after his death, Paul suggested that his work could be divided into two general categories: literary interpretations and textual criticism, with a focus on Middle High German, and the phonology and morphology of the Germanic languages. While Paul was certainly an important contributor to these fields, this self-classification is something of an oversimplification, as he made significant contributions to a number of other fields of linguistics and German studies.

Paul was certainly an influential figure in literary studies and textual criticism. His dissertation and *Habilitionsschrift* both dealt with questions of Middle High German literature. He authored a variety of articles on literary topics, ranging from the life of Hartman von Aue, an important Middle High German poet, to the question of whether a Middle High German literary language ever really existed. As to textual criticism, he was the founding editor of the *Altdeutsche Textbibliothek* [*Early German text library*], and edited three of the volumes of the series, all Middle High German texts. His *Mittelhochdeutsche Grammatik* [*Middle High German grammar*], first published in 1881 and republished many times (11 editions appeared in Paul's lifetime), remains a standard handbook in the field.

Paul's contributions to the study of Germanic phonology and morphology were also immense. He published a monograph-length article entitled 'Zur geschichte des germanischen vocalismus' ('On the history of germanic vocalism'), as well as articles on the vowels of the inflectional and derivational syllables in the oldest Germanic dialects and the ablative case in Germanic, among others. His most important work in this area is the five-volume *Deutsche Grammatik* (*German grammar*), described in his obituary as 'die einzige vollendete wissenschaftliche Grammatik der deutschen Sprache' ('the only completed scientific grammar of the German language').

Other areas of linguistics Paul contributed to include lexicography, syntax, word formation, and language pedagogy. He compiled a German dictionary, the *Deutsches Wörterbuch* (*German dictionary*), which remains in print today. Paul was also interested in the theoretical aspects of lexicography, publishing various essays on lexicographic topics (although the titles of these essays indicate that his main concern was with his German dictionary). As for syntax, Paul recognized its importance early on; his Middle High German grammar was one of the first such handbooks to include a section on syntax (although it must be admitted that this section was not in the first edition of the book; it was added to the 2nd edition, published in 1884). His work in word formation was generally an offshoot of his lexicographical work; he published an important article on word formation, 'Zur Wortbildungslehre' ('On the doctrine of word formation'), in 1896. Finally, a volume on language pedagogy, *Über Sprachunterricht* (*On language teaching*), was published posthumously in 1922.

Probably Paul's most significant contribution to linguistics was his work *Prinzipien der Sprachgeschichte* (*Principles of the history of language*), originally published in 1880. This volume rapidly became a standard handbook (five editions of the text were published during Paul's lifetime, along with an English translation of the second German edition), and proved to have a lasting influence on the field. It is notable not only for its incisive statements of many Neogrammarian ideas but also for its innovativeness, and is rewarding reading even today.

Finally, Paul's activities as an editor cannot be underestimated. Together with his friend and former fellow student Wilhelm Braune, Paul founded a journal, the *Beiträge zur Geschichte der deutschen Sprache und Literatur* [*Contributions on the history of the German language and [its] literature*], which is still generally referred to as the 'PBB', for *Paul-Braune Beiträge*. Paul served as coeditor of this journal for nearly 20 years, from 1874–1891, and was a consulting editor for 30 years after that. Paul was also the general editor of the *Grundriss der germanischen Philologie* (*Outline of Germanic philology*). Furthermore, as noted above, he was the founding editor of the *Altdeutsche Textbibliothek* (*Early German text library*), and edited three of the volumes of the series (listed above).

Despite his many contributions to the field, Paul is too often dismissed as 'merely' a historical linguist.

This viewpoint is entirely natural, since most of his work was on historical linguistics, and because of statements like the following, found in his *Prinzipien*: 'Es ist eingewendet, dass es noch eine wissenschaftliche Betrachtung der Sprache gäbe, als die geschichtliche. Ich muss das in Abrede stellen'. ['It has been claimed that there is another way to view language than the historical. I must object to this.'] On the other hand, it has also been claimed that Paul was much more than 'merely' a historical linguist. It has been argued that Paul was a forerunner of Structuralism; more specifically, that certain concepts found in the work of Ferdinand de Saussure, the well-known Swiss linguist who is generally viewed as the intellectual father of Structuralism (especially the distinctions between *synchrony* and *diachrony* and between *langue* and *parole*), were inspired by Paul's work (e.g. by his differentiation between individual linguistic acts and more general language use). The first of these viewpoints must be rejected; a closer reading of Paul's *Prinzipien* readily indicates that he was also interested in synchronic topics. The second of these viewpoints is probably also an overstatement. It is clear that de Saussure codified a number of ideas that were current linguistic theory at his time, but to trace the critical distinction between *synchrony* and *diachrony* to Paul, as some have, for example, is probably an exaggeration.

Regardless of these ideas, it is clear that Paul was an immensely important figure in linguistics. Paul played an extremely important role in the development of the Neogrammarian school of linguistics; he has been described by one historian of linguistics of the period as the Neogrammarians' 'most fertile and formidable theoretician', and his *Prinzipien* codified many aspects of Neogrammarian thought. More specifically, this same historian of linguistics (Kurt Jankowsky; the relevant work is cited under 'Further Reading' below) has argued that Paul's influence on his contemporaries can be traced by examining five major aspects of his work. First, Paul rejected the equation of linguistics to the natural sciences, suggesting that the term 'sound law' should not be understood as a 'law' is understood in physics and chemistry. Second, Paul distinguished between historical and comparative linguistics. Third, Paul was interested not only in *how* languages change, but *why* they do so. Fourth, Paul emphasized individual psychology over group psychology in his psychological approach to language. Fifth, Paul was deeply interested in linguistic geography, recognizing early on that a standard language is a convenient abstract fiction, made up of a variety of dialects and idiolects. A discussion of the place of the Neogrammarians in the history of linguistics is clearly beyond the scope of this entry, but their influence on linguistics, especially historical linguistics, was considerable. Numerous prominent linguists, in America and abroad, ranging from Leonard Bloomfield, the dean of American linguistics in the 1930s and 1940s, to William Labov, a prominent sociolinguist of today, were deeply influenced by Neogrammarian thought. Given the importance of the Neogrammarians, and Paul's importance within the group, his place in the history of linguistics seems assured.

Biography

Paul Hermann was born in Salbkempierc bei Magdeburg, Germany on, August 7, 1846. He attended the University of Berlin 1866–1867. He received a Doctorate (1870) for his dissertation on Freidank, Habilition (1872) for work on Gottfried's *Tristan*, University of Leipzig. He was Privatdozent in Leipzig in 1872–1874. He was also Professor of German, University of Freiburg im Breisgau, 1874–1893; Professor of German, University of Munich, 1893–1916; and Rektor, University of Munich, 1909–1910. He was a corresponding member of the Wissenschaftliche Akademie in Vienna in 1919–1921. Hermann died in Munich, Germany, on December 29, 1921.

References

Antal, Lázló. 1985. Some comments on the relationship between Paul and Saussure. Cahiers Ferdinand de Saussure 39.

Henne, Helmut, and Jörg Kilian (eds.) 1998. Hermann Paul: Sprachtheorie, Sprachgeschichte, Philologie. Tübingen: Niemeyer.

Jankowsky, Kurt R. 1972. The neogrammarians. A re-evaluation of their place in the development of linguistic science. The Hague: Mouton.

Koerner, E.F.K. 1972. Hermann Paul and synchronic linguistics. Lingua 29.

Labov, William. 1994. Principles of linguistic change. Internal factors. Cambridge, MA: Blackwell.

Paul, Herman. 1879. Zur geschichte des germanischen vocalismus. Beiträge zur Geschichte der deutschen Sprache und Literatur 6.

———. 1880. Prinzipien der Sprachgeschichte; transl. into English by H.A. Strong as Principles of the history of language, 1889; reprint of 8th edition 1970.

———. 1881. Mittelhochdeutsche Grammatik.

———. (ed.) 1894. Grundriss der germanischen Philologie.

———. 1896. Zur Wortbildungslehre. Sitzungsberichte der philos.-philol. Klasse der bayerischen Akademie der Wissenschaften.

———. 1897. Deutsches Wörterbuch.

———. 1916–1920. Deutsche Grammatik.

———. 1922. Über Sprachunterricht.

———. 1992. Mein Leben. Beiträge zur Geschichte der deutschen Sprache und Literatur 46.

Reis, Marga. 1978. Hermann Paul. Beiträge zur Geschichte der deutschen Sprache und Literatur 100.

Marc Pierce

See also **Historical Linguistics; Leskien, August; Saussure, Ferdinand de**

Peirce, Charles Sanders

Charles Sanders Peirce was a most original and highly versatile American philosopher, logician, and scientist—by all standards, a veritable polymath. Son of Benjamin Peirce, a distinguished Harvard mathematician, Charles was born in 1839 and grew up in a family milieu that was most conducive to a promising career in academia. Yet, except for a brief period as a lecturer in logic at the Johns Hopkins University (1879–1884), Peirce never did any systematic teaching. He was a difficult person to get along with and the university refrained from renewing his contract on charges of personal misconduct. His most sustained employment was with the U.S. Coast Survey, where he worked for 31 years and made a successful career as a geodesist, until the termination of his appointment in 1891. Thereafter, he set up a private practice as a chemical engineer and, thanks to the intervention of his friend and admirer William James, was invited to give occasional series of lectures at Harvard. His last years were spent in penury and failing health.

Peirce is widely known as the founding father of Pragmatism, a distinctively American philosophical movement whose central tenet is that belief guides action. Our ordinary beliefs as well as more sophisticated doctrines are nothing but dispositions to act in certain ways and are to be judged by taking into account their practical consequences. Peirce therefore rejected much of metaphysics—i.e. claims about the ultimate reality of what there is—since no practical consequences necessarily follow from alternative metaphysical conceptions. He was averse to the sort of intuitive rationalism as advocated by Descartes and foundationalism as encapsulated in the Cartesian quest for first principles, as well as his celebrated 'Method of Doubt'. As he famously put it in an essay published in 1878 titled 'How to make our ideas clear,' truth is simply that opinion which is destined to be agreed to by everyone who is disposed to finding it and so-called reality is made up of the objects represented in that opinion.

The idea of truth as resulting from a convergence of opinion was interpreted by many of his followers to mean that there is no more to it than a matter of mere consensus. William James for one regarded truth as 'whatever proves to be good in the way of belief'. But it is clear from Peirce's later writings that he subscribed to an objectivist conception of reality—the idea that there is a reality out there that is independent of what anyone believed about it. As for his epistemology, Peirce rejected the Cartesian attempt to use self-knowledge as the doorway to knowledge about the world, arguing instead that all knowledge is fallible, although continuous inquiry makes it self-corrective. He also contended that the ultimate reality is mental, matter being nothing but 'effete mind', and that science provides us with the best method of arriving at it.

Along with the Swiss linguist Ferdinand de Saussure, Peirce is also widely looked upon as a founding father of semiotics (or 'semeiotic', as he himself preferred to spell it). Unlike Saussure who claimed the sign relation to be dyadic—involving a sign (signifier) and a referent (signified) Peirce insisted that it was irreducibly triadic and involved, in addition to the object and the sign, an interpretant (the interpreting thought or meaning). The sign, in Peirce's conception, mediates between its object and its interpretant, which are therefore its correlatives. Central to Peirce's thinking is the idea that language does not simply and unproblematically refer to a reality external to it; it does so by *representing* it. This means, he claimed, all of cognition is necessarily mediated by the process of sign interpretation or semiosis. Peirce also viewed semiotics as fundamentally a classificatory science, like chemistry and biology. Furthermore, signs are not restricted to the activity of the human mind but are of the order of nature itself and, in this too, he differed from Saussure, for whom 'semiology' was part of social psychology. Unlike Saussure too, Peirce saw semiotics as teleological, tending toward the truth of things as they really are.

Peirce was a meticulous system-builder and his mind worked typically by positing trichotomies. Building on his own triadic characterization of sign relation, Peirce initially distinguished three divisions of signs: based on the nature of the sign itself, the relation between the sign and its object, and the role of the interpretant in representing the object. In their turn, these three divisions harked back to three ontological categories: quality (firstness), relation (secondness), and representation (thirdness). The first division, the one based on the nature of the sign itself, yielded a three-way distinction among 'qualisign' (appearance), 'sinsign' (token), and 'legisign' (type). Under the second division, the one obtained by taking into account the relation of the sign to its object,

Peirce distinguished icons (that bear a certain resemblance to the object), indices (that signify by direct causal interaction with the object in question), and symbols (that are associated with their object via convention). Finally, under the third division, the one where the interpretant enters the picture, yet another triad was postulated: rhemes (predicative signs), prepositional signs, and arguments. The arguments were further divided into the three categories—abduction (hypothesis/diagnosis), deduction (tautological reasoning), and induction (generalization)—the first two leading to qualitative or conceptual understanding and the last providing important quantitative details to it.

Toward the end of his life, Peirce grew increasingly impatient with the way the word 'pragmatism' was being used (and, in his view, abused) by other writers including William James and John Dewey and, in 1905, in a paper called 'What pragmatism is' published *The Monist*, pleaded that the word 'pragmaticism' be used instead to refer to his own doctrine, and distinguished it from that of the others, adding that the neologism was 'ugly enough to be safe from kidnappers'. There followed two other papers (forming a trilogy): 'Issues of pragmaticism' (1905) and 'Prologomena to an apology for pragmaticism' (1906).

Peirce died in 1914, an almost forgotten man. Early commentators often described him as 'a wasp in the bottle', someone who constantly came up against formidable odds and did everything but avail himself of the only way out of his troubles. Starting with the last quarter of the twentieth century or so, however, there has been a resurgence of interest in Peirce's work and a widespread recognition of the importance of his views to such areas as epistemology, linguistics, anthropology, psychology, and cognitive science.

Biography

Charles Sanders Peirce was born on September 10, 1839 in Cambridge, Massachusetts (USA). In 1855, he entered Harvard College, graduating in 1859 and continuing as a 'resident graduate' for one year. The same year, he began his career as a field aid at the Coast Survey, specializing in geodesy and working his way up until being appointed an Assistant. He entered Lawrence Scientific School at Harvard in 1861 and two years later obtained a graduate degree in Chemistry. He delivered Harvard lectures on 'The Logic of Science,' in the spring of 1865 and Lowell Institute lectures on 'The Logic of Science; or Induction and Hypothesis,' October 24,–December 1, 1866. He was elected to the American Academy of Arts and Sciences on January 30, 1867. He was Assistant at Harvard Observatory, October 1869–December 1872 and was elected to the National Academy of Sciences on April 20, 1877. He was appointed lecturer in Logic at Johns Hopkins in 1879 and was dismissed on charges of misconduct in 1884. He was elected to the American Association for the Advancement of Science in August 1880. He purchased a farm near Milford, Pennsylvania on May 10, 1888; over the next two years, he reconstructed the existing farmhouse and bought more land, and named the estate 'Arisbe'. The manuscript of 'How to reason' was rejected by both Macmillan and Ginn Co (1894) and that of 'New elements of mathematics' was rejected by Open Court (1895). In 1902, a grant application for 'Proposed memoirs on minute logic' was rejected by Carnegie Institution. He gave lectures on 'Pragmatism' at Harvard March 26–May 17, 1903 and Lowell lectures on 'Some topics of logic' November 23–December, 17, 1903. He was invited to deliver the Harvard Philosophy Club lectures on 'Logical methodeutic' April 8–13, 1907. Peirce died at Arisbe on April 19, 1914, of cancer.

References

Apel, Karl-Otto. 1981. Charles S. Peirce: from pragmatism to pragmaticism. English translation by John Michael Krois. New Jersey: Humanities Press.

Auspitz, Josiah Lee. 1994. The wasp leaves the bottle: Charles Sanders Peirce. The American Scholar 63.

Brent, Joseph. 1993. Charles Sanders Peirce: a life. Bloomington, Indiana: Indiana University Press.

Brock, J. E. 1974. Peirce's conception of semiotic. Semiotica 14.

Freeman, Eugene (ed.) 1983. The relevance of Charles Peirce. La Salle, IL: The Hegeler Institute.

Hookway, Christopher. 1985. Peirce. London: Routledge & Kegan Paul.

Ketner, Keneth Laine (editor) 1995. Peirce and contemporary thought. New York: Fordham Univerity Press.

Moore, Edward C. 1993. Charles S. Peirce and the philosophy of science. Tuscaloosa, London: The University of Alabama Press.

Murphey, Murray G. 1993. The development of Peirce's philosophy. Indianapolis, Indiana: Hackett Publishing Company Inc.

Peirce, Charles Sanders. 1931–1935. Collected papers of Charles Sanders Peirce, Vols. 1–6, ed. by Charles Hartshorne and Paul Weiss. Cambridge, MA: The Belknap Press of Harvard University Press; Vols. 7–8, ed. by Arthur Burkes. Cambridge, MA: The Belknap Press of Harvard University Press, 1958.

———. 1982. Writings of Charles S. Peirce: a chronological edition, 5 vols. Bloomington, IN: Indiana University Press.

Rosenthal, Sandra. 1994. Charles Peirce's pragmatic pluralism. New York: State University of New York Press.

KANAVILLIL RAJAGOPALAN

See also **Philosophy of Language; Sign Relationship**

Personality and Language

Personality and language are two terms that suggest some promising interconnections, such as the likely effect of personality pathologies on speech, the effect of personality traits on the use of certain linguistic—especially discursive—patterns, or the influence of individual learner differences on the acquisition of a second language. The effect of language on the configuration of personality is apparent if we think of the lexical terms used in a given culture to refer to personality phenomena. The ways in which people describe their own or others' personalities and the fundamental influence of a person's speech style on others' perception of his or her personality are examples of the interrelationship of the two. Additionally, personality may be seen as a key factor in the construction of language and of the speech style of each individual.

Notwithstanding these areas of inquiry, we are still far from claiming any conclusive direct effect of one over the other. Not surprisingly, it is speech (i.e. the oral performance of language by a given speaker) that has been the primary focus of researchers interested in the links between language and personality. These researchers have generally avoided claiming strong connections between personality and language as a system and have instead claimed a relationship between speech style and personality features, either self-assigned by the speaker or attributed by listeners.

Psychological research on personality has not been extremely consistent in providing a clear model against which to test any possible mutual influences of language and personality. Existing models range from viewing the structure of personality as composed of one or two dimensions, to the three or five factors that are currently the most widely accepted models. In the 1990s, the so-called Five Factor Model gained popularity within the research community. This model was first proposed in the 1960s, but it started receiving wide attention and empirical support three decades later. However, it still has to be confirmed in studies involving non-Western societies, where a different categorization of personality traits may be more accurate. In Western societies, the model seems to provide a good account of personality structure with the following five factors: extraversion, agreeableness, conscientiousness, emotional stability, and openness (also referred to as intellect).

The existence of a predominant model may eventually clarify research questions on the interplay between language and personality, and significant advances in this area may be seen in the near future. So far, the most intensive attempts to explore the common ground between the two areas of research have been conducted by social psychologists and—to a lesser extent—second-language researchers. Research has been primarily focused on two issues.

The first issue is the effects of linguistic differences on the attribution of personality features by listeners, whose evaluation is based solely on speech. These studies started with the seminal work of Wallace Lambert and his associates. They found that listeners in Montreal rated the personality of a given speaker differently when he spoke in two distinct languages (i.e. French and English) and also that French-speaking listeners tended to assign personality traits to the speaker quite differently from English-speaking listeners. Other speech evaluation studies of evaluations assigned to Spanish and English speakers in the United States followed. Howard Giles and his associates explored the effect of different regional and social speech varieties on personality ratings. Their research uncovered stereotypes linked to speech differences, and it was very valuable precisely in demonstrating that laypeople's connections between language and personality are the result of previously existing stereotypes. In all cases, two factors were identified: a competence factor, which had to do with efficiency and high status in society; and a solidarity factor, which involved a high affective empathy with the speaker. A common result of most studies conducted so far has been that language varieties that enjoy high prestige and can be regarded as the varieties of the most powerful people tend to be not so well valued in terms of personal empathy, and vice versa. More recently, some attempts have also been made to connect personality ratings and nonnative proficiency, with provisional outcomes that also point at the important role that language variation may play in listeners' perceptions of a speaker's personality, with all the major subsequent consequences for interpersonal relations. This method of study is less subjective than self-reports or self-accounts of personality traits, which tend to be heavily mediated by the idea of what one would like to be like rather than what one really is or seems to be. Another positive aspect of attributions of personality by listeners is the fact that they are closely related to social stereotyping. Thus, the study of the influence of

speech style on personality evaluation leads to the uncovering of social stereotypes.

The second area of research that links language and personality is second language acquisition, particularly the role of different personality traits in the ultimate achievement of second language proficiency. Surveys conducted among second language teachers and students show that personality is considered to be a primary factor in second language achievement. Empirical studies have mainly focused on a single personality factor: extraversion. Extroverted people seem to be more successful in acquiring basic communicative abilities because of their willingness to interact with other people and search for communication opportunities. Introverts are less engaged in communicative interactions but may benefit from an increased attention to the formal aspects of language. On the whole, research has not found one type of learner to be inherently better than another, but personality profiles seem to affect the preferred path to learning and, therefore, the final outcome of the process.

Personality and language should be viewed as two uniquely human characteristics that are bound to interact; however, at this point, there is a need for much more research to investigate the relationship between the two.

References

Ellis, Rod. 1994. The study of second language acquisition. Oxford: Oxford University Press.

Furnham, Adrian. 1990. Language and personality. Handbook of language and social psychology, ed. by Howard Giles and W. Peter Robinson. Chichester: John Wiley & sons.

Giles, Howard, and Peter F. Powesland. 1975. Speech style and social evaluation. London: Academic Press.

Goldberg, Lewis R. 1981. Language and individual differences: the search for universals in personality lexicons. Review of personality and social psychology, Vol. 2, ed. by L. W. Wheeler. Beverly Hills, CA: Sage.

Lambert, Wallace E., R.C. Hodgson, R.C. Gardner, and S. Fillenbaum. 1960. Evaluational reactions to spoken languages. Journal of Abnormal and Social Psychology 60. 44–51.

Llurda, Enric. 2000. Effects of intelligibility and speaking rate on judgements of non-native speakers' personalities. IRAL — International Review of Applied Linguistics 38. 289–99.

Ryan, Ellen B., and Miguel A. Carranza. 1975. Evaluative reactions of adolescents toward speakers of standard English and Mexican American accented English. Journal of Personality and Social Psychology 31. 855–63.

Saucier, Gerard, and Lewis R. Goldberg. 1996.The language of personality: lexical perspectives on the five-factor model. The five-factor model of personality; Theoretical perspectives, ed. by Jerry S. Wiggins. New York: Guilford.

Skehan, Peter. 1989. Individual differences in second-language learning. London: Edward Arnold.

Taylor, Harold C. 1934. Social agreement on personality traits as judged from speech. Journal of Social Psychology 5.

ENRIC LLURDA

See also **Discourse Analysis**

Philippine Spanish Creoles

The Spanish-based Creoles in the Philippines developed after the Spanish conquest in the sixteenth century and are commonly known as Chabacano. Despite its despicable original meaning ('tasteless', 'vulgar'), this name is used by the speech-community as a self-designation. Alongside Papiamento and Palenquero in the Americas, Chabacano is one of the three Spanish-based Creoles worldwide. It shares interesting features with other European-based Creoles in Asia, a fact differentiating it from the Atlantic creoles. Further, unlike the latter, Chabacano has been acquiring a mixed character; especially from the twentieth century onward.

The total number of Chabacano speakers can be estimated at no more than 500,000. Geolinguistically, Chabacano can be divided into two main varieties: Manila Bay Creole and Southern Mindanao Creole. Dialects of the former are found on the northern island of Luzon and are spoken by relatively small communities in the towns of Cavite and Ternate, but the dialect of Ermita, a district of Manila, can now be regarded as extinct. Southern Mindanao Creole is spoken on the island Mindanao in the south of the archipelago. The highest degree of vitality of Southern Mindanao Creole, and of Chabacano generally, can be observed in the city of Zamboanga and the surrounding area in the extreme western part of Mindanao (approximately 300,000 speakers). Creole is also used here as a lingua franca by diverse ethnolinguistic groups. It is codified for public and private purposes with a hispanic orthography; there are bible translations, literary writings, and until recently press. Public oral use has recently been increasing, there is broadcasting in the language, and it is used in church and partially in school. A subvariety of Zamboangueño is

also spoken in the town of Cotabato; however, little is known about the vitality of the subvariety of Southern Mindanao Creole found in Davao.

The genesis and history of Chabacano, especially Zamboangueño, is somewhat complex due to different contact situations in time and space. Keith Whinnom (1956) suggested that Chabacano grew out of the Portuguese creole of the Indonesian island of Ternate, which had been transported to the Philippines in the seventeenth century. He considered Zamboangueño a layer of Manila Bay Creole. This theory was contested principally by John Lipski (1992), who affirmed a typical creolization of a Spanish-based pidgin with Tagalog influences for Manila Bay Creole. For the special case of Zamboangueño, he proposes a development of different stages from the eighteenth century onward: it arose in the garrison, absorbing elements from Manila Bay Creole later on. During the nineteenth century, Hiligaynon elements were introduced through contact from ships coming from Manila having made a stop in Central Philippines. Afterward, it was re-hispanicized by native speakers from Europe and Mexico who came to Zamboanga. It was also influenced by Cebuano, like Hiligaynon, a Visayan language, which was brought by large-scale immigration. From the 1930s onward, the influence of English and, after World War II, that of the national language Pilipino have been increasing, especially affecting the lexicon. Consequently, Zamboangueño does not fit into the classification of plantation, fort, and maritime Creoles; nor can the dichotomy of exogene and endogene Creoles adequately describe the situation of Zamboangueño. In contrast to the Atlantic creoles, slavery played a marginal role. Whatever the different theories on the origin of Chabacano, first documented in Hugo Schuchardt's *Kreolische Studien* (1883; Creole studies), in the beginning there was a pidginized form of Spanish used for intercommunication by different ethnic groups, namely, European Spaniards, Mexicans, West-Austronesian speaking groups, and Chinese immigrants. A pidgin created by the latter is still spoken by shopkeepers in Davao ('Bamboo Spanish').

The current structure of the language, especially concerning the vocabulary and word structure, is heavily influenced by Philippine languages. The sound system shows Philippine features like the merging of [o]/[u] and [e]/[i] in nonaccented syllables and the occurrence of the glottal stop. Some of the productive morphemes for word formation in Zamboangueño originally come from Spanish, and many are from Hiligaynon or Cebuano (e.g. the verbalizing prefix *man-*: *kwénto* 'story' → *man-kwénto* 'to tell' or the adjectivizing prefix *ma-*: *pwérte* 'strength' → *ma-pwérte* 'strong'). Their form is often identical to those of the source languages, but in many cases they have different functions. The word-formation system seems to be unusually complex for a Creole. The plural of nouns is marked with the particle *maga* (*el maga péhro* 'the dog*s*'). In Zamboangueño, first-person plural pronouns have a Philippine form encoding inclusion or exclusion of the addressee. The basic word order of V(erb)-S(ubject)-O(bject) stems from Austronesian contact, e.g. *man-čángge yo* 'I'm going shopping'. This is exceptional compared to other European-based creoles that are typically SVO. As in Philippine languages, human proper names in subject function are marked with *si* (*ta-kantá si Maria* 'Maria is singing'). Human and definite objects are obligatorily marked with *kon*, perhaps a merger of an Austronesian element and Spanish *con* 'with', (*konosé-le kon ese muhér* 'he knows that women'). There are many discourse particles of Philippine origin (e.g. *daw* for reported information). However, Chabacano also has features typical of Atlantic creoles, e.g. serial verb constructions (*pwéde asé salé* 'to be able to make to go out'), no syntactic passive construction, lack of an equative copula (*soltéro el hénte* 'the man is a bachelor'), identity of 'have' and 'there is' (both *tyéne*), and preverbal markers for tense, mood, and aspect (e.g. *ya-andá yo na Samboanga* 'I went to Zamboanga', *ta-kusiná le* '(s)he is cooking' and *ay-kantá silá* 'they will sing'). The lexicon principally consists of Spanish vocabulary (83%), with some Philippine (c. 15%) and English (2.5%) words. It can be assumed that, more recently, the influence of English has increased and that of Spanish has diminished. Very few words are of Portuguese origin (e.g. *gumitá* < port. *gomitar* 'to vomit'). The importance of Mexican input is shown by words like *čangge* 'market' < *tiangue* (a Mexican word of Nahuatl origin). Despite the close geographical contact with Zamboanga, the local languages, Tausug, Subanon, and Yakan have nearly no influence on the lexicon.

References

Fernández, Mauro (ed.) 2001. Shedding light on the Chabacano language [Special Issue of Estudios de Sociolingüística. Linguas, sociedades e culturas 2, 27–56]. Vigo (Spain): University Press.

Forman, Michael L. 1972. Zamboangueño texts with grammatical analysis: a study of Philippine Creole Spanish. Ph.D. Dissertation. Cornell University.

Frake, Charles O. 1971. Lexical origins and semantic structure in Philippine Creole Spanish. Pidginization and creolization of languages, ed. by Dell Hymes, 223–42. Cambridge: Cambridge University Press.

Lipski, John M. 1992. New thoughts on the origins of Zamboangueño (PCS). Language Sciences 14. 197–231.

Papia. Revista de Crioulos de Base Ibérica 12 (2002) [Special Issue dedicated to Chabacano].

Riego de Dios, María. 1989. A composite dictionary of Philippine Creole Spanish (PCS). Studies in Philippine Linguistics 7. 1–210.

Schuchardt, Hugo. 1883. Kreolische Studien IV: über das Malaiospanische der Philippinen [Creole Studies IV: on the Malayo-Spanish of the Philippines]. Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften zu Wien 105. 111–50.

Steinkrüger, Patrick O. 2003. Morphological processes of word formation in Philippine Spanish Creole (Zamboangueño).

Phonology and morphology of Creole languages, ed. by Ingo Plag, 253–63. Tübingen: Niemeyer.

Whinnom, Keith. 1956. Spanish contact vernaculars in the Philippine Islands. Hong Kong: Hong Kong University Press.

PATRICK O. STEINKRÜGER

See also **Austronesian, Papiamento, Pidgins and Creoles, Spanish and Iberoromance Languages, Tagalog and Philippine Languages**

Philology

Etymologically, philology is the love of words (or, perhaps, the study of love). In the broadest sense, it covers everything having to do with the study of language—grammar, texts, history, civilization. Indeed, until a century or so ago, what is now called linguistics was simply part of philology, and even today, the term 'comparative philology' can still be found as a synonym for 'historical linguistics', especially within Indo-European studies. But here we should look at those branches of philology that have not been co-opted by (descriptive, historical, or theoretical) linguistics.

Thus narrowed, philology can be considered to be the study of texts, with all that implies: first, determining exactly what the author of a text actually wrote; then, determining what the author said; then, determining what the author meant. Correspondingly, a typical edition—whether of a book of the Bible, a play by Shakespeare, a novel by Joyce, or an inscription from an Indian temple—will contain the text itself, notes on every detail, and an interpretative commentary. Traditionally, the interpretative work of philology relies on the prior accomplishments of four subdisciplines: epigraphy, paleography, diplomatics, and textual criticism.

Epigraphy

Epigraphy is the study of inscriptions on hard material, such as stone and metal. People have been carving important texts into walls for as long as they have been writing—in Egypt, Mesopotamia, China, Anatolia, Greece, Arabia, Rome, Iran, India, Inner Asia, and Mesoamerica, in rough chronological order—and often such inscriptions are the only surviving written evidence of past civilizations. Sometimes they are official decrees, and by reading between the lines we can learn much about the societies that prompted them. Sometimes, as is almost exclusively the case with

Etruscan, they are funerary commemorations, from which we learn only of family relationships.

The task of the epigrapher is to record inscriptions accurately (and photography can rarely substitute for the sharp eye and steady hand of the draftsman), to publish them first promptly and then systematically, and to provide such tools as charts of letter forms that will assist in the interpretation of newly discovered inscriptions. A separate branch of epigraphy is numismatics, the study of coins and their inscriptions.

Paleography

Paleography is the study of manuscripts, written with pen or brush and ink on flexible material, namely papyrus, skin, or paper; textiles, wood, or leaves. These are perishable materials, so only rarely have they survived from more than a few centuries ago. The exceptions are due to extremely dry climates, as near the Nile and Jordan Rivers, areas where thousands of Egyptian papyri and the Dead Sea Scrolls have been found. In extremely rare cases, careful curatorial attention has preserved precious documents for a thousand years or more, but our knowledge of ancient texts results far more from a continual tradition of copying than from ancient attestations. (The tradition was fairly quickly superseded by printing beginning c. 1450.)

The paleographer's job is primarily to study the development of handwriting over generations of copyists, using dated manuscripts to anchor changes in stylistic trends and exercising judgment in assigning undated ones to this or that range of dates—although it must be recalled that any individual scribe's hand will not change much over their professional career, so a date based strictly on paleographical grounds cannot be more precise than within a generation or so. There are as many fields of paleography as there are scripts with manuscript traditions—cuneiform, Egyptian, Chinese,

Greek, Latin, Indian, Southeast Asian, etc. (there are no general surveys of world paleography outside histories of writing like Diringer (1968) and Jensen (1969); Urry (1974) considers only Greek and Latin).

Diplomatics

Diplomatics is the study of the form of documents. Its principal concern is detecting forgeries, so it deals principally with the physical properties of manuscripts and is the aspect of philology farthest removed from linguistics.

Textual Criticism

Textual criticism is the study of the history of the content of texts. Whenever a scribe makes a copy of a manuscript by hand, however carefully, mistakes are introduced. If a work was popular, many copies would be produced, and copies of copies, and so on, every one of them differing very slightly from all the others. By noting the correspondences among the differences in different copies, the chain of transmission can be reconstructed, and even if the author's original manuscript has not survived, it may be possible to reconstruct the author's original text with considerable certainty, with successive changes being removed.

Different scribal traditions handle the problem of scribal error in different ways. The rabbis who codified the text of the Hebrew Bible, known as Masoretes, did things like counting the number of words and even letters in each book. Scribes in the Islamic tradition included elaborate COLOPHONS in their manuscripts, naming not only the scribe and the date and circumstances of copying (as is often found elsewhere) but also a full list of all intervening scholar/copyists back to the original author. And Chinese scribes sometimes bypassed the problem by making copies of major works by printing them off woodblocks or even stone inscriptions of the texts, thus coming full circle in the fields of philology.

An interesting offshoot of textual criticism is that it is often possible to determine that a given text has been translated from another language, and that other language identified with considerable certainty. Thus, R.H. Charles was able to say of the Book of Enoch, which was known in full only in Ge'ez, with fragments surviving in Greek and Latin, that portions—chapters 1–5 and 37–104—were originally written in Hebrew, and chapters 6–36 were originally written in Aramaic (1912:lvii–lxx). Decades later, fragments of the Aramaic original were identified among the Dead Sea Scrolls.

Decipherment

The philological task most relevant to linguistics is the decipherment of disused scripts. Usually, these are brought to light by archeological excavations. Rarely, the unfamiliar inscription is accompanied by an inscription in a familiar script and language, and it can be assumed that the texts are at least roughly equivalent. This state of affairs is known as a BILINGUAL, and both the first decipherment, of the Aramaic language Palmyrene by Jean-Jacques Barthélemy in 1754, and the best known, of Egyptian hieroglyphics by Jean-François Champollion in 1822, involved bilinguals, although these two accomplishments were vastly different in difficulties. Pope (1999) includes detailed descriptions.

When no bilingual is available, the ingenuity of the philologist is required to discover a VIRTUAL BILINGUAL. The most challenging, and for its impact on our understanding of human history the most important, decipherment was of CUNEIFORM. Late-eighteenth-century travelers brought back to Europe drawings of inscriptions they found at the ruins of the capital of the Persian Empire, Persepolis, which was destroyed by Alexander in 330 BCE. Each inscription was in three different scripts, which were assumed to represent the same content in three languages (hence, they are called TRILINGUALS), but none of the scripts was familiar. The characters in all of them were composed of incised wedges (*cuneus* in Latin), whence the name.

The first success in interpreting cuneiform came in 1802. A German high school teacher, Georg Friedrich Grotefend, surmised that because the simplest of the three scripts always appeared in the most prominent position among the three, it probably wrote the ancient Persian language of the Empire. A few years earlier, the Frenchman Antoine Isaac Sylvestre de Sacy had deciphered royal inscriptions of a successor Iranian empire, the Sassanian, finding that they usually began with a genealogy in the form 'X, great king, son of Y, great king ….' Grotefend knew the names of the kings—Darius, Xerxes, and so on—from the Greek *Histories* of Herodotus and sought similar formulas at Persepolis. He soon found the expected recurring patterns in the inscriptions, and although he himself was not an Iranian philologist, his breakthrough discovery made it possible for specialists like the Danes Christian Lassen and Rasmus Rask to determine the basics of Old Persian. Edward Hincks, a Church of Ireland (Anglican) clergyman, showed that each of the 36 characters of Old Persian script represented a Consonant–Vowel (CV) syllable.

The second most complicated of the three cuneiform scripts had about 100 different characters (none of them the same as in Old Persian), so it was not surprising when, by comparing what seemed to be the same names in corresponding Persepolitan inscriptions, it too was shown to record syllables—both CV and VC. Its language, which is today called Elamite, is still not well understood. Hincks was one of the scholars who did that work; but more importantly, he went

on to decipher the third, most complicated script, which used several hundred characters. By the time Hincks got involved, in the mid-1840s, numerous cuneiform inscriptions had been brought from the heartland, Mesopotamia, and they too were in the third script, so Hincks had a great deal of material to compare. This time, the second script did constitute a subset of the third script, but the most useful aspect of the inscriptions was their repetitiveness: what seemed to be the same formulas appeared in different texts, or even within one text, but they contained small divergences. (Some of the best examples came from an extensive inscription in what we now call Urartian, which was published in 1840.) A sequence like *A B C D E F G* might substitute for *A B H E F G*, and Hincks saw that this could mean that *C D* stood for C_1V_1–V_1C_2, while *H* stood for $C_1V_1C_2$. He also discovered patterns like *iprus–iparras*, a sure sign that the language was Semitic; we now call it Akkadian (and the words were verbs). He found, rather incredibly, that characters could have more than one pronunciation—and he suggested that this was because the pronunciations found their origin in more than one language, what we now call Sumerian alongside Akkadian. Most importantly, he realized that some of the characters were being used not for the sounds they represented, but to stand for meanings. These characters—which he initially referred to as 'ethnological boulders,' since they were like alien rocks standing in a plowed field—were remnants of Sumerian routinely used in Akkadian texts.

Thus, between 1846 and 1852, Edward Hincks provided the key to the Akkadian, Urartian, and Sumerian languages. One often hears of H.C. Rawlinson in this connection, but he was never afterward able to explain how he deciphered cuneiform; it is now known that, in his post in Baghdad, he was kept informed of Hincks's work and used it in his analysis of the great inscription on a cliff at Behistun, Iran, which he copied with great difficulty—but which became available too late to be useful in the decipherment.

Another celebrated decipherment was accomplished by Michael Ventris in 1952, when he realized that the Linear B tablets from Crete and Mycaenae on the Greek mainland were actually in an archaic form of Greek, not some unknown predecessor language. In this and several other cases, the virtual bilinguals were known place names that seemed to occur in texts found in those places; another is the decipherment of the Maya glyphs.

The Uses of Philology

It is self-evident that without philology to make texts available for study, history could not be written and historical linguistics would be limited to deductions that can be made by comparing attested languages. But philology has applications that are not so obvious: investigating the *use and distribution* of languages themselves reveals patterns in the human career that do not emerge when only the *contents* of historical documents are considered. On at least one occasion a philologist has used his professional skills in creating a fantasy universe; Tom Shippey (2003), besides introducing the work of philology to the committed reader, shows how J.R.R. Tolkien's legion of imitators generally fail for lack of the kind of profound background Tolkien created through his many invented languages—and with example after example reveals that Tolkien's deep learning in Germanic philology, as well as Celtic and Finnic, underlies feature after feature of Middle Earth.

References

Charles, R. H. 1912. The book of Enoch …. Oxford: Clarendon.

Daniels, Peter T. 1994. Edward Hincks's decipherment of Mesopotamian cuneiform. The Edward Hincks bicentenary lectures, ed. by Kevin J. Cathcart. Dublin: University College, Department of Near Eastern Languages.

Daniels, Peter T. 1995. The decipherment of ancient Near Eastern scripts. Civilizations of the ancient Near East, ed. by Jack M. Sasson et al. New York: Scribner's.

Daniels, Peter T. 1996. Methods of decipherment. The World's writing systems, ed. by Peter T. Daniels and William Bright. New York: Oxford University Press.

Diringer, David. 1968. The alphabet: a key to the history of mankind, 3rd edition, 2 vols. New York: Funk & Wagnall's.

Herde, Peter. 1974. Diplomatics. Encyclopædia Britannica, 15th edition, Macropædia; since 1985 included under 'History, The study of.'

Jensen, Hans. 1969. Sign, symbol and script, 3rd edition, translated by George Unwin. London: George Allen & Unwin.

Kenney, E. J. 1974. Textual criticism. Encyclopedia Britannica, 15th edition, Macropædia; since 1985 included under 'History, The study of.'

Larsen, Mogens Trolle. 1996. The conquest of Assyria: excavations in an antique land 1840–1860. London: Routledge.

Ostler, Nicholas. 2005. Empire of the word: a language history of the World. London: HarperCollins.

Pope, Maurice. 1999. The story of decipherment: from Egyptian hieroglyphs to Maya script, 2nd edition. New York: Thames & Hudson.

Puhvel, Jaan. 1974. Epigraphy. Encyclopædia Britannica, 15th edition, Macropædia; since 1985 included under 'History, The study of.'

Robinson, Andrew. 2002. Lost languages:the enigma of the World's undeciphered scripts. New York: McGraw-Hill.

Shippey, Tom. 2003. The road to Middle Earth: how J. R. R. Tolkien created a new mythology. Boston: Houghton Mifflin; 1st edition, 1982.

Urry, William G. 1974. Paleography. Encyclopædia Britannica, 15th edition, Macropædia; since 1985 included under 'History, The study of.

PETER T. DANIELS

See also **Aramaic; Semitic Languages; Sumerian**

Philosophy of Language

Rather like the scholastic philosophers of the late Middle Ages, twentieth-century philosophers have seen philosophy as linguistic analysis, as the attempt to discern the logical structure of reality through discerning the formal structures—the superficial or deep grammar—of the language in which we report or think reality. Philosophers are particularly interested in certainty, in necessary truths as opposed to mere chance events. Nineteenth-century formalist mathematicians, similarly, honed a sense of mathematics as syntactically grounded in formal language or in a consistent set of linguistic conventions. Since philosophers conceived of their enterprise as a search for helpful logical truths and *not* as dependent on any experiential truths (philosophers do not do experiments), twentieth century philosophy became logicolinguistic analysis, and hence philosophy of language became, in short and for much of the past century, philosophy.

Linguistic philosophy was not just seeing the features of language that revealed the world's categorical structure but also seeing beyond language's perhaps misleading surface features. To give a standard example, the verb *to be* in English plays three vastly different logical roles. *Is* can mean *identity,* as in 2 *is* (*the same as*) 2, *Dubya is* (*the same as*) *President George W. Bush,* or *heat is* (*the same as*) *the average motion of molecular particles. Is,* however, can also mean predication, as in *The sky is (has the property of being) blue, Roses are* (*have the property of being*) *red,* or *The earth is* (*has the property of being*) *spherical.* Finally, *Is* can mean existence, as in *There is* (*exists*) *a Santa Claus* or *God is* (*exists*) as opposed to *God is (exists) not.* Bertrand Russell (1919), while imprisoned as an antiwar demonstrator in World War I, wrote in *Introduction to mathematical philosophy* that it was important to keep these three senses wholly separate in linguistic analysis (so important that Russell added that he would declaim so even if he were 'dead from the waist down and not merely in prison'). In the mathematical logical notation that Gottlob Frege and Russell created in the late nineteenth century, these notions were indeed wholly separate and specifically represented. With a notation that puts predicates in capitals and writes names and variables for individuals in lower case, 'a = b' means two names stand for the same individual, 'Pa' means predicating P of individual a, and '∃x[Dx]' means, if D predicates *divine*, that there exists a God. Russell stressed that this notation commendably dissolved the 'ontological argument for the existence of God', which mostly simply runs, *God is that being who possesses all possible perfections; existence is a perfection; therefore, God exists.* This argument confounds the hypothetical predications 'If something is Divine, then it has various Perfections' and the existential 'Something exists that is Divine'.

Russell was particularly praised for his 'paradigmatic' analysis of 'definite descriptions', phrases of the form *The so and so.* Consider 'The present King of France is bald.' If you think this statement has a subject/predicate logical form, and believe that statements are either true or false, then you seem to have to say that *either* 'The present King of France is bald' is true *or* 'The present King of France is not bald' is true. The problem, of course, is that there is no *present King of France* (when Russell wrote and subsequently). It is no solution to say 'The present King of France' means *nothing,* because then 'The present King of the United States' and 'The round square' and countless other phrases would presumably also mean, or stand for, *nothing;* but it is evident that all such phrases differ in meaning. If a revolution occurred in the United States in 2084, there might then be a King of the United States, but that would not mean there would be a round square and a French King as well (or that *nothing* had changed to *something*). Russell insists that 'The present King of France' does not *mean anything* all by itself. Rather, 'The present King of France is bald' means 'There exists an *x* that bears the predicate of *being kingly of France*, and if any *y* also bears that predicate, *y = x*, and *x* is bald.'

Following Russell's lead, in the 1920s, the logical positivists of the Vienna Circle insisted that logico-mathematical truths provide the formal structure within which the observational truths of experience array themselves. *Theoretical terms,* such as 'vital spirit', 'electron', or 'virus', are only acceptable if they can be specified completely in observational terms. The Vienna Circle philosopher W.V.O. Quine (1960) of Harvard University claims that by casting science into Russell's austere predicate logic, you can most clearly determine what science says has to be real. Indeed, such a translation into predicate logic will strip off the possibly misleading superficial features of actual languages. Philosophical *linguistic analysis,* conceived in this fashion, cares nothing for the *phonology* of language, for literal physical sound streams and their transformation into the sharper, leaner, and deeper

structures that fluent speakers hear. Moreover, the exclusive concern is with the truth or falsity of sentences that describe the world, collectively feeding what scientific generalizations we can muster about the world. As Quine put it, startlingly, *all of science is held up as one sentence to nature*—but that is just to insist that we collectively assert and understand this scientific sentential consensus, not that anyone actually *says* this *one sentence*. Philosophy of language, as so understood, has little concern with *how natural language is used in the acts and interactions of everyday life.* Quine makes this clear when, at the beginning of *Word and Object*, he approvingly cites Otto Neurath's metaphorical remark that we are like seamen who must make repairs in our boat while at sea. Here, 'we' means all scientific-minded humans, and 'the boat' means whatever natural and artificial symbol systems we use in our collective scientific description of the world, with Quine highlighting the austere notation of mathematical logic.

By the mid-twentieth century, however, philosophers began to shift from a concern with the true/false logical relationship between sentences and the world to a more expansive, but also more narrowly linguistic, concern with sentences as speech actions of speakers, who carry individual responsibility *to us* for what they say, indeed for what they *do* in saying *so*. In the 1950s, the Oxford University philosopher J.L. Austin, among others, drew attention to *performatives*, sentences that, when uttered by the appropriate person in the appropriate circumstances, *do something.* If I say, 'I promise to return your $20 tomorrow', I am not describing, truly or falsely, some peculiar mental state; rather, by my speaking, *I make a promise.* If I say, 'By the power vested in me by the State of New Jersey, I appoint you Port Commissioner,' *I make an appointment.* 'I promise X' and 'I appoint Z' are *explicit* performatives, where the main verb of the sentence explicitly indicates what action is performed. However, 'I certainly will return your $20 tomorrow' will, given appropriate circumstances, constitute a promise. Similarly, if an officer says to his subordinate, 'You will move your men to that bridge,' he has given an *order.* Indeed, Austin maintained that every *use of language*, or every speech act, has a performative aspect. For example, if I say 'There's a bittern in your garden,' I should know a little something about bitterns and have had some opportunity to identify the bird. If I have no idea of what bitterns look or sound like, and indeed, have not been in or near your garden, I *have no right to say* what I did. Given that I meet those minimal requirements, I may demur if you ask, 'How do you know it's a bittern?' pleading perhaps 'Well, I don't *know* it's a bittern but it is a large, white-feathered marsh bird', or I may take a further plunge and say, 'Oh, I know it's a bittern—I got a clear view and, growing up in the Fens, I'd know that booming anywhere'. Philosophical concern with semantic, performative, and pragmatic aspects of language has meant some fruitful interaction with linguistic science, particularly given the concern with syntax and logical form stressed by Noam Chomsky and other generative linguists since the 1960s.

But still other philosophers have come to feel that philosophy is well rid of an exclusive emphasis on philosophy as linguistic analysis. In the middle decades of the twentieth century, philosophers concerned with values emphasized metaethics or the 'logic of the language of morals'; more recently, philosophers have addressed specific normative issues. Similarly, many recent philosophers have constructed rational choice, and social contract theories. Thinkers as diverse as the logician Saul Kripke and the linguist Noam Chomsky have argued that necessary truths are more central to science than the real but often trivial analytic truths of language. *Heat is the average motion of molecular particles* is not, Kripke argues, an analytic or lexical truth, but rather a scientifically discovered necessary physical truth. In the same vein, *water is H_2O* asserts a physical necessity, given the basic combinatory properties that form the fabric of atomic physics. Similarly, perhaps, the linguist's claim that natural languages are generative and transformational follows from the structural character of the human linguistic faculty as a matter of biological necessity (it certainly does not follow from a lexical entry attributed to normal speakers, most of whom do not have 'generative and transformational' in their everyday vocabulary—or 'average motion of molecular particles', for that matter). In any case, few philosophers today would be wholly content to characterize philosophy as *linguistic analysis.*

References

Austin, J. L. 1961. Philosophical papers. Oxford: Clarendon Press.

Ayer, A. J. 1936. Language, truth, and logic. London: Gollancz.

Chomsky, Noam. 1986. Knowledge of language. New York: Praeger.

Kripke, Saul. 1972. Naming and necessity, in Semantics of natural language, ed. by Gilbert Harman and Donald Davidson, Dordrecht, Holland: D. Reidel.

Quine, W. V. O. 1960. Word and object. Cambridge, MA: MIT Press.

Russell, Bertrand. 1919. Introduction to mathematical philosophy. London: Allen & Unwin.

JUSTIN LEIBER

See also **Austin, John Langshaw; Chomsky, Noam; Quine, Willard van Orman; Russell, Bertrand**

Phoneme

The phoneme [from Greek *phonema*—a sound] is recognized as the smallest abstract sound unit of a language, for example, *map* consists of three phonemes /m/, /æ/, and /p/. Each word in a given language must consist of a permitted sequence of phonemes. Phonemes may be defined as the class of sounds of a language that distinguish words from one another.

The phoneme is a fundamental unit of phonology—the study of the sound systems of languages and the relationships between those sounds. Different languages have various numbers of phonemes; for instance, one Brazilian language—*Piraha*—is said to have only ten phonemes, while one of the African languages—*!Xu*—is known to have 141 phonemes.

The term phoneme was introduced in the late 1870s, and the phoneme theory was first developed by Jan Baudouin de Courtenay and his pupils as a response to the need to systematize the sound patterns of the language. Since then, the phoneme and its features have been defined in various ways by different linguists. As Baudouin understood it, the term originally meant a 'mental image' of a real physical sound, and speech was the act of producing sounds as close to that mental image as possible. However, during the twentieth century, linguists came to a different view. Lev Scerba, who had initially followed Baudouin, shifted to the functional aspect of the phoneme, that is differentiating words. Daniel Jones also described a phoneme as a physical concept. If Nikolai Trubetskoy defined the phoneme as the smallest distinct unit and incapable of further subdivision, Roman Jakobson saw it as the sum of features that distinguish the phonemes from each other.

Contrastive Function

It is traditionally accepted that the phoneme functions as the minimal contrastive unit of speech. In other words, the phoneme differentiates the meaning of words. The native English speaker easily discerns the difference in meaning between the words in minimal pairs, i.e. word pairs that differ in only one sound element: *man* and *pan*, where [m] and [p] differentiate between a human and a kitchen utensil, thus indicating that they possess different phonemes; *bag* and *big*, where [æ] and [I] make possible the distinction between an object and a size; *den* and *then*, where [d] and [ð] clearly indicate that one word means a 'lair' and the other one means 'at that time'; or *man* and *men*, where [æ] and [e] identify one man as opposed to two or more of them.

Phonetic Context

Every uttered word is a particular combination of sounds and is called the phonetic context of a sound. The English phonetician Daniel Jones gives the following definition of the phonetic context:

> …phonetic context of a sound is to be understood to mean the sounds next to it or near it in the sequence of which it is a part, together with its duration (length), stress and (if voiced) voice-pitch.

Minimal pairs provide the same phonetic context with only one different phoneme, as in *bat* and *cat*; where the sounds [b] and [k] occur in the same phonetic context, they differentiate the meaning of the words, and, therefore, are two different phonemes.

Allophones

A phoneme may have several variants or *allophones*. As opposed to phonemes, allophones do not distinguish words, and they never occur in the same phonetic context. For example, *cat* and *king* possess a common sound [k]. However, the sound [k] is pronounced slightly differently: in the first instance, it is articulated—or pronounced—further back in the throat. In the second instance, the tongue—in an attempt to combine the sounds [k] and [i]—moves upward and forward producing a palatalized [k]. These two sounds vary only in the way they are articulated (pronounced); they cannot occur in the same phonetic environment, i.e. the palatalized [k] cannot occur before [æ] as in *cat* for the simple reason that our tongues are physically unable to make these two sounds together. The two sounds vary slightly in articulation; however, the difference is not significant, and may not even be picked up by a native English speaker. Thus, these two variants of /k/ are not distinctive in English; they are not different phonemes, but rather allophones of the same phoneme /k/.

Daniel Jones gives the following examples of the allophones of the English phoneme /t/:

The principal [t] is used before vowels in the strongly stressed position as in *table* or *ten*.

A dental [t] is used before [th] as in eighth.
A retracted [t] is used in [tr] as in train.
An unaspirated or barely aspirated [t] is used in weakly stressed positions as in *letter.*
A laterally exploded [t] is used in [tl] as in *kettle.*
A nasally exploded [t] is used in [tn] as in *mutton.*

For a native language speaker of a given language, it is not really important how carefully allophones are pronounced to understand what is being said. However, it is more important that phonemes are pronounced properly in order for the listener to get the correct message; instead of *My brother bought you a big present*, the speaker must not relay *My brother bought you a pig present*.

Languages differ with respect to which sounds are perceived as distinctive (phonemic) and which sounds are mere variants or allophones of phonemes. English distinguishes between [l] and [r] as in *lay* and *ray*, which make them different English phonemes. Korean does not see any distinction between these two sounds, i.e. they are perceived by Koreans as allophones of one phoneme. Moreover, although English distinguishes between [d] and [ð] as in *than*, Malay does not distinguish them as two phonemes but as allophones of one phoneme. Of course, many languages make other distinctions that an English native speaker may not be able to hear.

Distribution

Each language imposes limitations on the distribution of phonemes, i.e. each particular phoneme can occur only in certain phonetic contexts in that language. In English, there are no words that can start with the nasal [ŋ] sound. There are no such words as *ngull, ngreed, ngoose*. In Chinese or Vietnamese, however, similar sound constellations are quite possible (e.g. names such as *Ng* or *Nguyen*).

The distribution pattern of the phonemes in a language can be identified by analyzing which phonemes may occur in the following positions:

syllable-initial   syllable-medial   syllable-final
word-initial   word-medial   word-final

The English phoneme /p/ can be found in all of the above-mentioned positions, for example:

post   apt   map
packet   upper   ketchup

Other English phonemes can occur only in certain positions, for example, /æ/ as in *rat, stamp, sachet, ant* can only be found in syllable- or word-initial positions and syllable- or word-medial positions.

Phonological Opposition

The phonological opposition, developed by the Prague School, illustrates the network of connections and relations between the phonemes. For example, *proportional* opposition features two or more pairs of phonemes, such as /p/ vs. /b/, /k/ vs. /g/, which are opposed to each other on the basis of voiced features, while /r/ and /l/ are in *isolated* opposition and are opposed to each other on features that do not create a phonemic opposition in English.

Bilateral opposition represents only pairs of phonemes, such as voiced–voiceless pairs /z/ - /s/, where the third element is not possible. *Multilateral* opposition, on the other hand, may have three or more elements, such as /p/, /t/, and /k/, which are all stop consonants, or /f/, /ð/, /θ/, /s/, which are all fricative consonants.

Privative opposition can be illustrated by voiced–voiceless pairs, in which one member of each pair lacks a voiced feature (/d/-/t/, /b/-/p/). *Equipollent* opposition is where voiceless /p/, /t/, /k/ each have a different place of articulation. *Gradual* opposition involves gradation of a feature in the phonemes, for example, in vowels /i/, /e/, /a/, the difference is in aperture, or close–open continuum of tongue height.

In many languages, the phonological opposition can be neutralized. For example, in English, when /s/ occurs in the word-final position, it is pronounced as [s] after a voiceless plosive: *lips, rats, sacks,* and as [z] after the voiced plosive: *dogs, bags, ribs.*

In the Russian language, a voiced consonant can never occur at the end of the word. When the voiced consonant is found at the word-final position, it must be pronounced as voiceless. Neutralization in Russian can be illustrated by the following pair: *pok* spelled as [rok] (type of music), and *poг* spelled as [rog] (a horn); both are pronounced as [rok]. In the latter instance, the voiced [g] sound is neutralized and is pronounced as the voiceless [k].

Major Early Works on the Phoneme: Baudouin de Courtenay, Lev Scerba, Ferdinand de Saussure, Daniel Jones

This section is a brief excursion into the major contributions of world linguists in relation to the notion of the phoneme only, and not phonology, which is beyond the scope of this paper.

The term phoneme was introduced in 1875 to indicate the basic unit of human speech. Although it is said that the term was first used by French linguist Dufriche-Desgenettes in 1873, it was Jan Baudouin de Courtenay and Mikolaj Kruszewsky of the University of Kazan (Russia) who, between 1875 and 1895, conducted an extensive study of sounds, and established the phoneme theory. The Kazan School focused its research on the phoneme as 'the psychological equivalent of the speech sound'. The phoneme, according to Baudouin, was impressed in the speaker's mind. In the speech process, the speaker was aiming at uttering a sound similar to the psychological impression or the mental image; however, the pronunciation was always determined by the real phonetic context.

The work of Baudouin influenced the Russian linguist Lev Scerba (1880–1944), who attended Baudouin's

lectures in St. Petersburg. Scerba accepted the psychological definition of the phoneme, but later concentrated on the functional aspect of the phoneme, i.e. the differentiation of words. Scerba also introduced such phonological notions as complimentary distribution, which is a characteristic of a sound that never occurs in the same phonetic context. This can be illustrated by allophones of a phoneme, all of which occur in different phonetic contexts but are variants of the same phoneme. Since the allophones are different realizations of the same phoneme, it can be said that they are in complementary distribution to each other.

Ferdinand de Saussure (1857–1913), a Swiss linguist, is considered the founder of modern linguistics. His main conception assumed that language is a system of mutually defining entities. He believed that language was a structured system. This view had a major impact on the later development of structuralism in linguistics, which studies the language as a system and not the individual elements of it. Structuralism was further developed in Europe by Trubetskoy and Jakobson, and in America by Sapir and Bloomfield.

Daniel Jones (1881–1967), who was a lecturer of phonetics at University College, London, favored the idea of the phoneme as a practical tool for teaching purposes. His contribution to linguistics was in developing a more practical approach to phonology. He found phonology very useful for acquiring good pronunciation by people learning a foreign language, as a basis of nonphonetic branches of linguistics such as morphology and grammar, and as an important tool for creating alphabets as simply as possible.

The Prague Linguistic Circle, which later came to be known as the Prague School, began its meetings in 1926. It consisted mainly of Russian and Czech linguists. Among them were the Russian linguists Nikolai Trubetskoy and Roman Jakobson. The contribution of Trubetskoy lies in the developing of the notion of phonological opposition, and relations between the members of an opposition. Jakobson's view of the phoneme as a sum of distinctive features led to the creation of a new approach to phonological description. This approach of distinctive features of

the phoneme was based on the structuralist ideas of de Saussure. For example, a sound can be a vowel (a, i, o, e, u) or a consonant (b, d, g, z). If the sound is a consonant, it can be voiced (b, g, d) or voiceless (p, k, t), nasal (m, n) or nonnasal (all others). *Vowel, voiced,* and *nasal* are all examples of distinctive features.

American structuralism was developed by American linguists Edward Sapir and Leonard Bloomfield. Sapir adhered to the idea that a native speaker knows the phonemic system of their language through intuition. For Sapir, the phonemic system was a mental reality that existed independently of the act of speech. Bloomfield saw the phoneme as reality, a bundle of distinctive features, but not as the intuition of a speaker. He believed that phonemic analysis was the process of isolation of its distinctive features through minimal pairs. According to Bloomfield, finding a distribution of redundant, nondistinctive features was also part of the phonemic analysis.

References

Abercrombie, David. 1965. What is a letter? Studies in phonetics and linguistics. London: Oxford University Press.
———. 1967. Elements of general phonetics. Edinburg: University Press.
Bloomfield, Leonard. 1933. Language. New York: Holt.
Hawkins, Peter. 1984. Introducing phonology. London: Routledge.
Jakobson, Roman. 1962. Selected writings, Vol.1. The Hague: Mouton.
Jones, Daniel. 1962. The phoneme: its nature and use, 30. Cambridge: W. Heffer & Sons Ltd.
Kramsky, J. 1974. The phoneme. Munich: Wilhelm Fink Verlag.
Trubetzkoy, Nikolai Sergeevich. 1958. Grundzuge der Phonogie, Gottingen: Vandenhoeck & Ruprecht; third edition, 1962, as Principles of phonology, translated by C.A.M. Baltaxe, Berkeley: University of California Press, 1969.
Yallop, Colin. 1995. English phonology. Sydney: Macquarie University.

ALFIA ABAZOVA

See also **Baudouin de Courtenay, Jan Ignacy Niecisław; Bloomfield, Leonard; Feature Theory; Sapir, Edward; Saussure, Ferdinand de; Structuralism; Trubetzkoy, Nikolai Sergeyevich**

Phonetic Transcription

In speech, different letter(s) may represent the same sound (e.g. s*ee*, gr*ie*f, k*ey*) or the same letter may represent different sounds (e.g. b*a*d, m*a*ny, w*a*nt). In order to be able to express speech sounds of each

language known in the world in an unambiguous way, the International Phonetic Association developed a set of symbols known as The International Phonetic Alphabet (IPA). The use of sequences of phonetic

symbols to represent speech is known as phonetic transcription. The concept of a phonetic alphabet is to have one symbol for one sound. Over the years, different alphabets with modified IPA symbols have been developed for specific aims (as many symbols are not used), but the IPA is considered the standard.

History

In 1886, the International Phonetic Association was founded in Paris by language teachers who wanted phonetic notation to be used in schools, as a method of acquiring a realistic pronunciation of foreign languages. In 1897, it was named L'Association Phonétique des Professeurs de Langues Vivantes or the International Phonetic Association in English. The IPA is used, among other things, to indicate pronunciation in a dictionary, to record a language, and to transcribe speech for automatic speech recognition. The IPA is based on the Roman alphabet, but other symbols have been included to cover the wide variety of sounds found in the languages of the world. Phonetic transcription is placed between square brackets [ ] and there is no capitalization or punctuation. An overview of the different characters and symbols of the IPA is given in various charts elsewhere. The symbols will be discussed in more detail after highlighting some of the principles underlying the phonetic alphabet.

Phonetic transcription does not only vary from language to language but also from speaker to speaker. Depending on dialect, fluency, etc., the same word can be transcribed differently. Speech segments can be divided into two major categories: consonants and vowels. They are described with reference to how they are produced and their auditory characteristics. It requires considerable skill to transcribe an unknown language. During speech production, the vocal tract varies continuously and the division between subsequent speech sounds may not be clear.

Only linguistically relevant speech sounds are transcribed, not personal voice quality. A separate sign exists for each distinctive sound (e.g. /k/ for cat, track, kite, quick, monarch). The IPA also provides symbols for suprasegmental information (word boundaries, stress, intonation) and for refining the pronunciation of an utterance (e.g. aspiration).

IPA charts

The IPA charts give an overview of the different vowels, consonants, clicks, and other sounds occurring in the languages of the world. The charts are more than a list of symbols; they show how the different types of sounds can be classified. Consonants

can be distinguished with regard to voicing, manner, and place of articulation. Voicing refers to the presence or absence of vocal fold vibration (/b/ versus /p/, /d/ vs. /t/). Manner of articulation refers to the way in which speech sounds are produced (stop, fricative, nasal, trills, etc.), and place of articulation indicates the place of constriction in the vocal tract (lips, alveolar ridge, velum, pharynx, etc.). Some languages in the world contain nonpulmonic sounds, sounds that are not produced through air from the lungs: clicks and voiced implosives are indicated by special symbols, while ejectives are indicated by a plosive and an apostrophe. From an articulatory point of view, vowels can be classified according to the features 'open', 'close', 'front', and 'back'. A vowel is considered 'open' when the space between the tongue and the roof of the mouth is large and 'closed' when the tongue is near the roof of the mouth. In front vowels, the tongue is fronted and raised toward the alveolar ridge; in back vowels, the tongue is near the back of the mouth (palate). Although the vowel space is continuous, most vowels are placed in relation to one of the eight cardinal (=reference) vowels: /i, e, ɛ, a, ɑ, ɔ, o, ʊ/.

Several symbols are foreseen to designate suprasegmental information, i.e. intonation groups (‖), duration (ː), stress ('), and to distinguish between words in tone languages (e.g. Chinese, Thai). Some of these symbols are iconic in the sense that the shape of the line indicates the height and possible movement of pitch. For example, [maꜜ] with falling intonation means 'to scold' and [maꜛ] with fixed high intonation means 'mother' in standard Chinese. However, it is also possible to transcribe pitch height by adding diacritics to existing segments, e.g. [ó bà] meaning 'it perched' vs. [ó ba] 'he or she hid' in the West African language Yoruba. In this transcription, the symbols are not iconic, i.e. the accent does not indicate a rising or falling pitch, it means 'high' 'mid', or 'low'.

Another set of marks, called diacritics, has been included to further refine the transcription of phonation. This has been done to restrict the total number of characters. Diacritics are used to indicate aspiration, e.g. [pʰaɪ] 'pie', to indicate vowel centralization [ë] relative to the cardinal vowels, breathy and/or creaky voices, nasalization [fɛ̃] 'fin', meaning 'end' in French, and many other features.

Broad and Narrow Phonetic Transcription

Speech sounds can be transcribed 'broadly' or 'narrowly'. A broad transcription is a phonemic transcription, while a narrow transcription captures more phonetic details of the speech sounds. For instance, the word 'toast' can be transcribed phonemically as

[təʊst] or phonetically as [tʰɵ̆ʊst]. Both types of transcription require some knowledge of the phonology of the language to solve ambiguities related to the segmentation of utterances. And even when the likeliness of occurrence of certain (combinations of) phonemes is known, the alignment between the physical speech sounds and the phonetic transcription can still be problematic. Phonetic transcription has proved to be a very useful skill, but it cannot be done automatically. A proper analysis of the speech signals requires considerable knowledge about the segmental and suprasegmental aspects of a language.

References

Albright, R.W. 1958. The International Phonetic Alphabet: its background and development. International Journal of American Linguistics 24(1). 3.

Handbook of the International Phonetic Association. 1999. A guide to the use of the International Phonetic Alphabet, 204pp. Cambridge: Cambridge University Press.

Ladefoged, P. 1982. A course in phonetics, 2nd edition. New York: Harcourt Brace and Jovanovich.

Ladefoged, P., and Maddieson, I. 1996. The sounds of the world's languages, 426pp. Oxford: Blackwell Publishers.

Roach, P. 1991. English phonetics and phonology, 262pp. Cambridge: Cambridge University Press.

ASTRID VAN WIERINGEN

Phonetics

Phonetics is the scientific study of speech. It is concerned with all aspects of the production, acoustics, and perception of speech in the languages and dialects of the world. The starting point for almost any phonetic investigation is the identification of certain landmarks from a cross-sectional view of the vocal tract from which both consonants and vowels of the languages of the world can be classified (see the section on Basic Classificatory Principles). The classificatory system can be used to describe both the various types of speech sounds that occur in different languages and their dialects, as well as those aspects of pronunciation that are used contrastively, that is, to signal differences of meaning (see the section on Phoneme and Phonetic Variation). Phoneticians also analyze how languages make use of pitch, duration, and loudness to communicate meaning distinctions at both the level of the word and the utterance (see the section on Prosody).

Phonetics is concerned with the way in which spoken communication is accomplished between speakers and hearers. This includes both the relationship between speech sounds and the neurological, physiological, and neuromuscular aspects of speech (see the section on Speech Motor Control and Physiology), as well as the resulting acoustic signal that is decoded by the listener (see the section on Speech Acoustics and Speech Perception).

Phonetics is a highly interdisciplinary field and is informed by theory and methodology from Linguistics, Cognitive Science, Computer Science, Electronics, Signal Processing, Acoustics, Neurology, Anatomy, and Physiology. Within Linguistics, Phonetics overlaps most directly with Phonology, Psycholinguistics, Sociolinguistics, and Dialectology and also with Syntax, Semantics, and Pragmatics in modeling the prosodic aspects of speech.

Phonetics is relevant to disorders of speech communication and to speech pathology. In the last quarter of the twentieth century, phonetics has contributed in various ways to the field of speech technology, which is concerned with the development of computer systems for the automatic generation (synthesis) and recognition of speech.

Basic Classificatory Principles

The range of possible sounds that can occur in the world's languages and from which a language makes a selection can be described in terms of various interactions between the vocal organs (Figure 1). The great majority of speech sounds are produced when air is expelled from the lungs, passing through the larynx between two shelves of muscular tissue known as the vocal folds (Figure 2). When the vocal folds are apart, as they are in quiet breathing, they do not obstruct the passage of air from the lungs. Many speech sounds are produced with open vocal folds and they are known as voiceless. In many other speech sounds, including all vowels, the vocal folds can be drawn together and made to vibrate very rapidly, that is, they repeatedly alternate between being open and closed—such sounds are voiced. There is a direct relationship between the rate at which the vocal folds vibrate and the sound's pitch. In most languages, there are pairs of
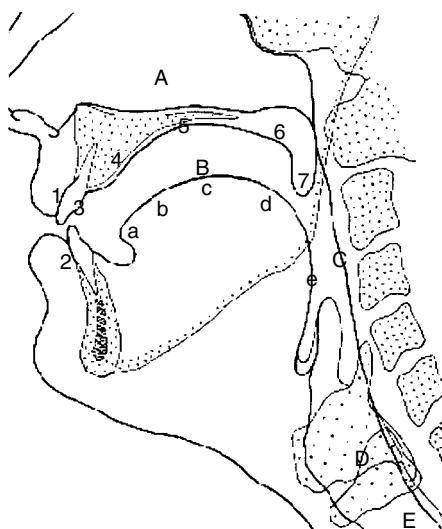
Figure 1. A cross-sectional view of the vocal tract of an adult male taken from an X-ray image (adapted from Laver, J, 1994). Upper lip (1), lower lip (2), upper front teeth (3), alveolar ridge (4), hard palate (5), velum or soft-palate (6), uvula (7), tongue tip (a), tongue blade (b), tongue front (c), tongue back (d), tongue root (e), nasal cavity (A), oral cavity (B), pharyngeal cavity (C), larynx (D), and trachea (E). The places of articulation are obtained from the following combinations: bilabial (1+2), labiodental (2+3), dental (a+3), alveolar (a+4), apical postalveolar (a+5), laminal postalveolar (b+5), palatal (c+5), velar (d+6), uvular (d+7) and pharyngeal (e+C).
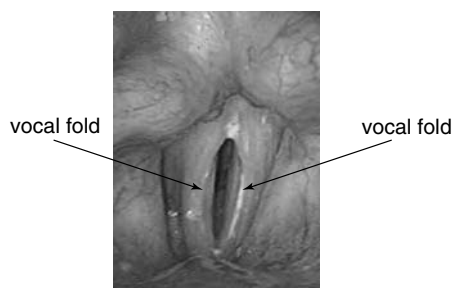
Figure 2. A photograph of the vocal folds taken from above (from Ladefoged, 2001b). The space between the vocal folds is called the glottis. Here the vocal folds are apart allowing the air to pass from the lungs through the trachea into the rest of the vocal tract.

sounds that differ only in whether or not the vocal folds vibrate, such as [f] (in 'fan', voiceless) and [v] (in 'van', voiced). Voiced sounds are usually accompanied by vibrations that can be felt by placing a hand at the level of the throat and then switching voicing on and off as in a repeated production of [fffvvvfffvvvf-fvvv].

The air from the lungs can be made to pass out of the vocal tract either through the mouth (the oral cavity) or the nose (the nasal cavity) depending on the
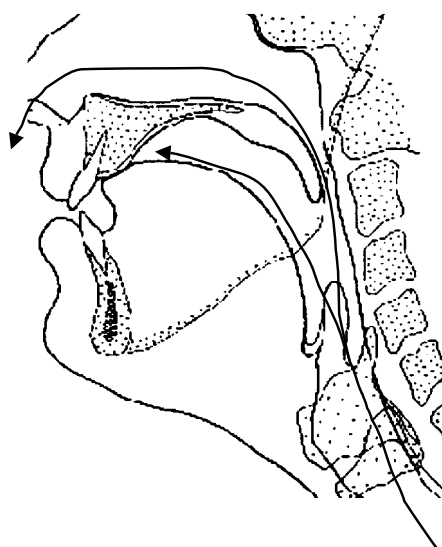
Figure 3. The configuration of the vocal organs for the production of [n]. The direction of the air from the lungs is shown by the lines and arrows. The velum is lowered allowing air to enter the nasal cavity and pass out through the nose.

actions of the velum or soft palate (Figure 3). When the velum is raised, the air passes out of the mouth or oral cavity and sounds produced in this way are said to be oral. When the velum is lowered, the air enters and passes through the nasal cavity and exits through the nose: such sounds are nasal. Many languages have pairs of sounds that are distinguished only by whether or not the velum is raised. For example, in English [b] in 'bad' and [m] in 'mad' are produced in the same way, except that [b] is oral and produced with a raised velum while [m] is nasal and produced with a lowered velum.

Within the oral cavity, there are three other important landmarks: the upper front teeth, and behind them the alveolar ridge, which extends in an arch into the hard-palate (Figure 1).

The tongue is a highly flexible and mobile vocal organ that has a very dense concentration of different muscles. The tongue is subdivided into the tip, blade, front, and back; in the rest position, these four landmarks lie roughly below the alveolar ridge, the back of the alveolar ridge, the hard palate, and velum, respectively. The tongue is attached to the epiglottis just above the vocal folds and includes a near-vertical section in the pharynx, the root.

The upper and lower lips are involved in various ways in speech production: they may be protruded as in the vowel of 'soon' or completely closed as in [b] and [m] in '*b*ad' and '*m*ad'.

In producing any consonant, there is a point in the vocal tract at which there is the greatest degree of narrowing and therefore the greatest obstruction to the

flow of air: its location is the consonant's place of articulation and it can vary anywhere between the lips and vocal folds (Figure 1). Three commonly occurring places of articulation are: bilabial (e.g. [p], [b], [m] in 'pan', 'ban', 'man'), **alveolar** (e.g. [t], [d], [n], [l] in 'tie', 'die', 'nine', 'lie), and **velar** ([k], [g], [ŋ] as in 'cot', 'got', 'sing').

For any place of articulation, there can be variations in stricture, which defines the extent to which the flow of air out of the mouth is obstructed. When the obstruction is complete, as in [b d g p t k m n ŋ] in the above examples, the consonant is phonetically a stop. A fricative is produced when the air is forced through a very narrow opening at high speed, often producing a hissing sound due to the air becoming turbulent: examples are [s] ('sip') and [ʃ] ('ship'). If the obstruction is further reduced, the flow of air is no longer turbulent but laminar. A consonant produced in this way is an approximant; examples in English include [w] ('we'), [j] ('you'), [l] ('led') and in some English dialects [ɹ] ('red'). [l] in 'led' is also a lateral, in which the air flows over the side of the tongue, and is distinguished from central sounds in which the air flows over the center of the tongue.

The two principal features for describing vowel production are height and backness. Height is analogous to stricture and defines the degree of opening in the mouth. The smallest opening, which is never less than for approximants, occurs in high or close vowels as in 'heed' while the greatest opening is in low or open vowels like 'had'. Backness is analogous to the place of articulation defined earlier: vowels can vary between front, in which the narrowing occurs at the hard palate, and back in which the greatest point of narrowing is often in the pharynx.

Height and backness together form a two-dimensional space that defines the ranges within which vowel production (Figure 4) occurs. The edges of the vowel space are marked by the cardinal vowels, defined by the phonetician Daniel Jones in 1918 at the University
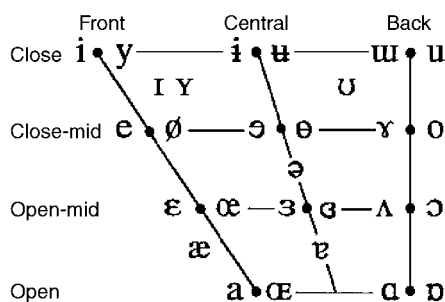
College of London. The Height × Backness space is used to define the phonetic differences between vowels of different languages and dialects. There is independent experimental evidence to show that the extent to which listeners judge two vowels to be similar depends to a large extent on the distance between them in this space. Various acoustic analyses in the last 40 years have shown that height and backness are correlated, respectively, with the first two resonances or formant frequencies of the vocal tract.

Vowels can also vary in whether they are rounded (produced with protruded lips) or unrounded (the lips are unprotruded). There is a preference across languages for front vowels to be unrounded (e.g. 'heed', 'head', 'had' have unrounded front vowels) and back vowels to be rounded (e.g. 'hoard', 'who' have rounded back vowels). But there are many languages with front rounded (French, German) and back unrounded vowels (Japanese, Vietnamese) and in some languages, the presence or absence of lip-rounding may be the only feature that distinguishes between vowels (e.g. in French: high, front, unrounded [vi] ('vie', 'life') vs. high, front, rounded [vy] ('vu', 'seen')).

The International Phonetic Alphabet provides a set of symbols for transcribing the sounds of the world's languages based on classificatory labels or features of the kind discussed above. Each symbol therefore provides information both about the sound that is produced and its relationship with other sounds. For example, [p] is a voiceless bilabial oral stop, which means that like [f] and [t], it has open vocal folds (voiceless), like [m] and [b] it has a constriction at the lips (bilabial), like [b], [f], and [t] it has a raised velum (oral), and like [b], [m], [t], and [d] the passage of air through the mouth is completely obstructed at some stage during its production (stop).

The set of features and their relationships with each other have been considerably informed by a detailed instrumental analysis of the sounds of the world's languages in the last 40 years (e.g. Ladefoged and Maddieson 1996). Other research seeks to understand why certain feature combinations tend to be preferred by languages, whether babies are born with detectors for identifying features, and the extent to which listeners make use of features in perceiving speech.

Phoneme and Phonetic Variation

The consonants at the beginning of the words 'pin' and 'bin' are different from each other not only phonetically (the first is voiceless, the second voiced) but also in a more abstract phonological sense: they communicate a difference of meaning and are allophones of separate phonemes. The two 'p' sounds in 'pin' and 'spin' are also phonetically very different from each

Figure 4. The vowels and their phonetic symbols of the international phonetic alphabet. The vertical dimension is Height, the horizontal dimension is Backness. When the symbols appear in pairs, the one to the right represents a rounded vowel.

other in most accents of English: the one in 'pin', transcribed as [pʰ], is aspirated and produced with a puff of air that can be felt by saying the word and holding a hand near the lips, whereas the one in 'spin', transcribed as [p], is unaspirated and is produced without any puff of air. But while [pʰ] and [p] are phonetically very different, they do not communicate a difference in meaning and are allophones of the same phoneme /p/. The fact that these sounds do not distinguish between meanings in English can be shown by saying the word 'spin' but with an aspirated [pʰ]—the pronunciation might be unusual, but native speakers of English would nevertheless still identify the word 'spin'. Allophones of the same phoneme occur predictably in different contexts: whereas [p] occurs in the context after /s/ ('spin', 'spoke', 'spring', etc.), [pʰ] occurs in a different context at the beginning of stressed syllables ('pin', 'pan', 'put', etc.).

A type of transcription that makes use of only the phonemes of a dialect or language is known as a broad transcription: in such a transcription, all the phonetic characteristics that are predictable from context, and, therefore, redundant for distinguishing between meaning, are filtered out. A narrow transcription is one that includes some phonetic detail, usually at least all the allophones that are common to the speakers of a particular dialect.

Every language has a finite number of phonemes out of which it assembles its words to construct differences of meaning. An important goal in phonetics and phonology is to establish both the phoneme inventory of a language and the relationships between the phonemes and their allophones. All of these differ markedly between languages and dialects. For example, while [p] and [pʰ] are allophones of the same phoneme in English, in Korean they are allophones of separate phonemes: as a result, exchanging them can produce a meaning difference in Korean but not English. The same phonemic opposition is nevertheless likely to give rise to phonetic differences between two languages. For example, both English and French create differences of meaning by distinguishing between two phonemes /b/ and /p/ ('bin' vs. 'pin' in English; 'beau' ('beautiful') vs. 'peau' ('skin') in French). But in French the corresponding phonetic distinction is between a fully voiced [b] (the vocal folds vibrate while the lips are closed) and an unaspirated [p], whereas in English the distinction is between a devoiced [b̥] (the vocal folds only start vibrating after the lips come apart) and an aspirated [pʰ].

Not only languages but dialects of the same language can differ in their phoneme inventory and how these phonemes are produced, or realized phonetically. Whereas most English dialects have a phonemic difference between /u:/ in 'pool' and /ʊ/ in 'pull', in

Scots English there is no such phonemic opposition and so these words are indistinguishable. And while most English dialects have two very different allophones of /l/ depending on its position in the syllable, Southern Irish English makes no such distinction. Therefore, /l/ in 'leaf' and 'feel' are phonetically very similar in Southern Irish English but quite different in most other dialects.

Phonetic variation occurs at the level of the individual speaker. This variation is not simply due to the anatomical differences between speakers (which is responsible for considerable acoustic differences in the speech of men, women, and children) but also because speakers may have developed an idiosyncratic speaking style: a given speaker might have excessive nasalization or a tendency to protrude the lips during speech production. Developing a model that can represent speaker-specific aspects of pronunciation is part of voice quality research and it has important applications both in disorders of voice and speech as well as in the forensic analysis of speech that is increasingly used in criminal investigations.

Prosody

The same utterance can be said in many different ways to communicate paralinguistic effects of emotion such as happiness, sadness, surprise, anger, and so on. These paralinguistic effects are usually brought about by manipulating the prosody or the timing, pitch, and loudness of an utterance. But these manipulations can be used linguistically to distinguish either between words (word-level prosody) or between utterances (sentence- or utterance-level prosody). Prosodic effects usually extend over, and modify, more than a single consonant or vowel; this is why the units of prosody are sometimes called suprasegmentals.

Word-level prosody includes differences due to quantity that are largely communicated by timing differences (these are marked phonetically by a colon: thus, [t:] is a long version of [t]). Quantity differences can occur in the consonant (e.g. Italian: [fato], 'fact' vs. [fat:o], 'done') or in the vowel (German: [lam] ('lamb') vs.[la:m] ('lame')) or in both (Finnish: [mut:a] ('but') vs. [mu:ta] ('other') vs. [mu:t:a] ('to change')). Word-level prosody also includes tone differences in which word distinctions are based on pitch: for example, in Thai, the syllable [na] has different word meanings depending on whether it is produced with a high-falling pitch (when it means 'face'), a low-falling pitch (when it is a name), or a high-rising pitch (when it means 'aunt')—see Ladefoged (2001a) for examples and sound files.

The syllable and word-stress are part of word-level prosody. Almost all languages organize their phonemes

into syllable units and there is psycholinguistic evidence to show that children can identify syllables from a very early age. Much of the phonetic variation (such as the two types of /l/ discussed above) can be attributed to the structural position of a sound in the syllable. In many languages, consonants at the end of a syllable are phonetically much weaker than at the beginning of a syllable. This is the source of many sound changes in which syllable-final consonants weaken or delete completely (e.g. Latin 'septem' > French [set] ('sept', 'seven'), processes that also occur synchronically (e.g. the /t/ of 'fast' is perhaps never pronounced in a context like 'fast speech', so that 'fast' rhymes with 'pass').

Word-stress has to do with the relative prominence or salience and, to a certain extent, the relative loudness of syllables: for example, in the word 'abracadabra', the first and fourth syllables are more salient than the others and are said to be stressed, while the others are unstressed. In English, the position in the word of the syllable with the strongest word-stress, known as the primary stressed syllable, can vary (e.g. '*patt*ern' vs. 'ad*mit*'), whereas in other languages like Icelandic and Polish it has a delimitative function and occurs at a fixed position in the word. In some languages, there are a handful of words that differ only in their word-stress pattern: German 'übersetzen' means 'to transport' (e.g. by ferry) when it has primary stress on the first syllable but 'to translate' when it has primary stress on the third syllable. In English, unstressed syllables (and the sounds within them) can be considerably shortened and often completely deleted (e.g. the second unstressed syllable in 'Tower Bridge' can be deleted resulting in a pronunciation close to 'Tar Bridge'). This very clear phonetic difference between stressed and unstressed syllables that affects vowels and consonants in English is much less is evidence in languages such as French and this is one of the attributes that contributes to the very differently sounding rhythm of these two languages.

Utterance-level prosody includes intonation and accentuation, both of which can be used linguistically to provide a range of different meanings. Intonation depends phonetically on the rise and fall in pitch and can mark syntactic differences in many languages such as distinguishing a statement from a question. Languages, and indeed dialects of a language, have quite different associations between meaning and intonation. For example, whereas in most accents of English, the intonation falls toward the end of a statement, in Belfast-English it usually reaches a plateau and may rise; and whereas English, in common with many languages, has a rising intonation in questions requiring a 'yes—no' answer ('Did Marianna make the marmalade?'), the intonation of such 'yes–no' questions in Greek, Hungarian, and Romanian usually falls.

In the same way that word-stress concerns the relative prominence of syllables in a word, accentuation has to do with the relative prominence of words in an utterance. In many, but by no means all, languages, shifting the accentuation can evoke a different meaning that depends on the discourse structure of a dialogue. Consider that when said in isolation, the most prominent or accented word of the sentence 'I don't like classical music' is usually 'music'. But if the same sentence were said in reply to: 'Have you ever listened to Beethoven?' then the accent in the same reply would shift to 'like' to indicate that 'classical music' is part of the background or old information in the dialogue that is shared between the speaker and listener.

Accented words tend to be produced with greater clarity precisely because they often carry new information and are therefore much more difficult to predict from the context in which they occur. For a related reason, function words, which rarely carry new information and which are therefore much less likely to be accented, are less clearly produced and may be entirely deleted: consider that 'the man in the moon' can be understood as such simply by lengthening the [n] of 'man' and deleting the vowel in 'in' entirely.

Speaking style can have a considerable influence on the utterance's prosody. At faster rates of speech, or in a more casual conversational setting, many of the types of weakening and deletion discussed in this section are likely to be in greater evidence.

Speech Motor Control and Physiology

Speech motor control is concerned with the neurological and physiological bases of speech production. Investigations in this area are directly relevant to articulatory synthesis in which the aim is to be able to synthesize speech using a model of the speech production mechanism.

Studies of speech motor control make extensive use of instrumentation for the analysis of the vocal organs including: laryngography and direct fiberoptic laryngoscopy for measuring and viewing the activity of the vocal folds; aerodynamic techniques for measuring airflow and air pressure at various points in the vocal tract; and electropalatography, which records the pattern of contact between the tongue and the roof of the mouth. Midsaggital electromagnetic articulometry (EMA) has been in use for about 15 years for recording the movement and velocity of the jaw, the lips and the tongue; this technique has to a certain extent replaced electromyography, which measures the electrical activity associated with muscle contraction. Considerable progress has recently been made with magnetic resonance imaging (MRI), which provides exceptionally clear cross-sectional images of the vocal

tract, and with ultrasonic techniques, which have been used for recording the movement of the vocal folds and the tongue based on measuring reflections of high-frequency sound waves between tissue and air.

A major issue in speech motor control is understanding a type of phonetic variation known as coarticulation. This comes about because, in contrast to the way in which letters are written on a page, sounds in sequence overlap with, and therefore influence, each other. Consider for example the production of 'pan': the last two phonemes /an/ are oral–nasal meaning that the velum is raised for the /a/ vowel and lowered for /n/. But since the velum can not jump instantaneously between these two states, it already starts to lower during the /a/ making the vowel nasalized well before the tongue tip touches the alveolar ridge for /n/. The production of /a/ is therefore influenced by the following /n/: it is because of this type of coarticulation that the /a/ in 'pan' is phonetically very different from the /a/ in other contexts such as 'pad'. Coarticulation is ubiquitous and, since speech sounds can be communicated at a faster rate if they overlap with each other in time, proponents of the motor theory of speech perception (see next section) have argued that coarticulation is necessary to ensure that speech is produced sufficiently rapidly. Coarticulation may be anticipatory as in the 'pan' example, in which a sound is influenced by a following sound, or perseverative in which the influence is from a preceding sound. Various experiments have shown that it is very difficult to tie coarticulation definitively to any category boundary: coarticulatory influences can spread across phoneme, syllable, and word boundaries; hence, it would be quite possible for the velum to start lowering during the vowel of 'saw' in anticipation of the following /n/ in 'saw another'. Explaining how listeners manage to recover the phonemic and linguistic content of an utterance intended by the listener in the light of these ubiquitous coarticulatory effects is a major research undertaking in speech motor control and its relationship with speech perception.

Another major issue is the extent to which speech motor control is regulated by feedback—that is, information that is relayed back to the higher centers of the brain from hearing oneself speaking (auditory feedback) and detecting the movement of, and contact between the vocal organs (tactile and proprioceptive feedback). Research in this area has been informed both by studying the way in which fine motor control in adults who have become deaf in later life deteriorates, and from so-called immediate compensation experiments, in which investigators seek to determine the extent to which a vocal organ like the jaw can compensate for the lack of movement in the lips, if these are artificially immobilized. Recent research

suggests that feedback may be used to develop an internalized model of feedback: this would allow speech production control to be a good deal more rapid because, instead of having to wait for actual feedback to determine whether speech has been produced as intended, speakers could calculate the feedback they would be expected to get and make any necessary corrections before speech was actually produced.

Speech Acoustics and Speech Perception

This branch of phonetics is concerned with the analysis of acoustic speech signals. Major progress in this area has been possible as a result of at least three major technological developments: the invention of the spectrograph in 1948, allowing an acoustic signal to be represented in terms of its time, frequency, and amplitude components; progress in speech synthesis technology in the 1950s and 1960s, in which utterances could be generated synthetically by machine; and advances in digital speech processing in the 1960s, in which computer algorithms were developed for the rapid calculation of the frequency content of a signal and for representing acoustic speech signals in terms of a small number of parameters.

When a speech sound is produced, an acoustic signal is created whose characteristics are entirely dependent on the actions of the vocal tract that give rise to it. In articulatory-to-acoustic relationships, the aim is to break down the acoustic signal into a number of components that can be related to individual vocal tract actions out of which the speech sounds are formed: we would like to know, for example, how the acoustic signal is changed by opening and closing the velum, protruding the lips for a rounded vowel like [u], opening the mouth for [a], and so on. A major breakthrough in our understanding of such relationships was in the development in the 1950s and 1960s of the acoustic theory of speech production. With such a model, it became possible to reduce the complex shape of the vocal tract to a relatively small number of parameters (for example, vowels can be modeled by the cross-sectional areas and lengths of four interconnecting cylinders) and then to use these parameters to predict the likely acoustic output. This type of model, which was developed primarily by Gunnar Fant in the 1950s and 1960s has been central to a range of research areas in phonetics, including synthesizing speech and predicting the types and distribution of vowels and consonants that are most likely to occur in the world's languages.

When an acoustic signal reaches the listener, it is transformed in various ways as it passes through the ear and is represented by electrical impulses in the

auditory nerve. Auditory phonetics seeks to model these types of transformations in order to obtain a more accurate representation of how listeners actually perceive an acoustic signal. Some researchers hold the view that the auditory transformations reduce the considerable acoustic differences between male and female speech.

Speech perception is concerned with how listeners retrieve phonemes and linguistic units from the acoustic signal. The application of speech synthesis to speech perception—allowing the individual contribution of acoustic cues to the perception of speech to be investigated—was pioneered at the Haskins Laboratories in the United States of America in the early 1950s. Their experiments led to the influential motor theory of speech perception, in which it is proposed that listeners extract or decode phonemes from the acoustic signal by first reconstructing the speech production strategies that could have given rise to them: that is, listeners hear not the acoustic signal but the movements of the vocal tract. In Kenneth Stevens' lexical-access from features model by contrast, which is partly based on the well-known quantal theory of speech perception, it is proposed that there are landmarks of acoustic stability that allow listeners to access the mental lexicon directly without recourse to speech production. In Björn Lindblom's hyper and hypo (H&H) model of speech, the notion that listeners make use of either invariant acoustic or articulatory landmarks is rejected. Instead, speech is said to vary along a continuum from hypo- to hyperarticulation, depending on the speaker's desire to produce speech with a minimum of effort (hypoarticulation) and the need to produce speech clearly (hyperarticulation) at e.g. points of accent and information focus (see above).

Speech perception is further complicated by the way in which decoding linguistic information from the speech signal interacts with 'top-down' processing, that is, with the knowledge of the language that the listener brings to bear in recognizing and understanding speech. It is quite clear that listeners do not recognize speech by first decoding the speech signal into phonemes and subsequently transforming these into words: a listener instead often recognizes a word well before the acoustic signal for that word has occurred in its entirety. Listeners can identify words that may be present only in a substantially impoverished form in the acoustic signal, or perhaps not even physically present at all (see the examples above). A further complicating factor is that, in contrast to the presence of white spaces between printed words on the page, there is no direct information in the acoustic signal about where one word ends and the next begins. Various experiments have suggested that listeners may make

use of prosody (in particular, syllable-structure and word-stress) to divide the stream of speech into separate words. Developing a model of how the available information from the acoustic signal interacts with the listener's knowledge of the language is a major challenge both to understanding how speech is perceived by humans and to developing computers to accomplish an analogous task.

Conclusions

Phoneticians have made considerable progress in understanding how abstract phonemic and prosodic units are related to the physical characteristics of speech sounds in the production and perception of speech in the languages of the world. The acoustic theory of speech production and the synthesis of intelligible speech from text are two examples of significant scientific discoveries in the last 50 years that have very clear practical applications in technology.

The dramatic improvement in the storage and analysis of speech data by computer in the last decade has provided the tools for analyzing a far greater range of speaking styles of naturally occurring spontaneous speech. However, most of the phonetic knowledge is still derived from stylized laboratory speech of English and there is a great lack of comprehensively annotated databases from other languages and in particular from endangered languages. New experimental approaches to measuring the vocal organs have advanced our understanding of speech production control. There is now recorded material spanning several decades, which can be used to provide an experimental basis to sound change. Careful analyses of all these new types of data will lead to more sophisticated models and new discoveries in phonetics in the future.

References

Clark, J., and C. Yallop. 1995. An introduction to phonetics and phonology, 2nd edition. Oxford: Blackwell.

Fant, G. 1973. Speech sounds and features. Cambridge, MA: MIT Press.

Johnson, K. 1997. Acoustic and auditory phonetics. Cambridge, MA: Blackwells.

Ladefoged, P. 2001a. A course in phonetics, 4th edition. New York: Harcourt Brace College Publishers.

Ladefoged, P. 2001b. Vowels and consonants: an introduction to the sounds of language. Oxford: Blackwell.

Ladefoged, P. 2002. Phonetic data analysis: an introduction to fieldwork and instrumental techniques. Oxford: Blackwell.

Ladefoged, P., and I. Maddieson. 1996. The sounds of the World's languages. Oxford: Blackwell.

Ladd, D. 1996. Intonational phonology. Cambridge: Cambridge University Press.

Laver, J. 1994. Principles of phonetics. Cambridge: Cambridge University Press.

Laver, J., and W.J. Hardcastle. 1997. The handbook of phonetic sciences. Oxford: Blackwell.

Lehiste, I. 1967. Readings in acoustic phonetics. Cambridge, Ma: MIT Press.

———. 1970. Suprasegmentals. Cambridge, MA: MIT Press.

Pickett, J. 1998. The acoustics of speech communication. Boston: Allyn and Bacon.

JONATHAN HARRINGTON

See also **Spectral Analysis**

Phonology

The aim of phonology is to examine the way sounds are organized in languages and to explain the variations that occur. While it is physically possible to produce a wide range of sounds, only a relatively small number of these are used in a language. For example, a nasal preceding a plosive, such as in *mbeke* does not occur in English at the beginning of a word, but it does in several African languages. This is not because English speakers cannot produce the specific sequence of sounds, but because of the way speech sounds in a particular language are organized.

In phonology, the most characteristic properties of different sounds, known as features, are compared to develop rules underlying the use of sounds in groups of languages. Systematic surveys of a representative number of languages are necessary to be able to generalize about sound systems, and to relate the findings to other areas of language (e.g. syntax and morphology). The UPSID database (**U**CLA **P**honological **S**egment **I**nventory **D**atabase), developed by Ian Maddieson and colleagues in the 1980s, contains inventories of over 450 languages of the world. This database, as well as the Stanford Phonology Archive, has become a popular tool in teaching and in phonological and phonetic research. A detailed description of a language is not only of interest for the sake of having a survey of different languages of the world but is also essential to be able to teach foreign learners a language, or to treat children who are not acquiring the sound system of their language properly, or people who have lost the capacity to speak through injury or illness. In addition, speech technology communication systems require knowledge of the structure of a language.

Phonemes and Allophones

A classical approach in phonology is to begin by establishing the phonemic system of a language, i.e. to determine which sounds are phonemes and which are allophones. Substitution of one phoneme for another will result in a word with a different meaning; substitution of one allophone for another one only results in a different pronunciation of the same word. Phonemes are the 'contrastive' sounds of a language: /c/ and /s/ are different phonemes, because 'cat' and 'sat' contrast in meaning. Allophones are the phonetic variations of distinctively used sounds. Sounds that are not distinctive are also known as 'redundant'. In English, aspiration is a redundant feature. However, in Thai, aspiration is distinctive, as the use of aspiration affects the meaning of the utterance. By substituting one sound for the other (the 'commutation test'), it is possible to determine the phonemes of a language. Pairs of words that differ by just one sound, such as 'cat' and 'sat' or 'robe' and 'roam', are called minimal pairs.

Other criteria also help to determine whether speech sounds belong to the same phoneme or not. One is complementary distribution, which refers to the situation where two sounds should not occur in the same environment. For instance, in English, aspirated plosives occur at the beginning of words, but not in consonant clusters. It can be predicted which allophone is pronounced in which context. Where we find the one, we do not find the other: they are mutually exclusive, never occurring in the same phonetic environment. Another criterion is free variation: sounds that occur in the same place in a word can belong to the same phoneme only if they do not change the meaning of the word. For instance, substitution of glottal stop [ʔ] for [t] as in 'butter' does not change the meaning of the word. Therefore, /buʔer/ and /butter/ are not minimal pairs. Also, sounds ought to display a reasonable amount of physical similarity to belong to the same phoneme.

In summary, when two sounds are in complementary distribution or free variation, they are allophones of the same phoneme.

Distinctive Features

A phoneme can be described by several articulatory and acoustical features. One of the main aims of

phonology is to identify the set of distinctive features required to describe the sounds of a language. Distinctive features serve to distinguish one phoneme from another. It is important to describe these detailed aspects of speech sounds, in order to understand how sets of sounds are related.

Tables 1 and 2 list some distinctive features (not exhaustive) of some English vowels and consonants. If a feature is present, it is marked with a (+) sign; if it is not present, it is marked with a minus (−) sign.

The features 'front', 'back', 'high', 'low', and 'round' refer to the position of the tongue in the vocal tract. For front sounds, the body of the tongue is fronted *re* its neutral position (for the /ə/, a schwa), for back sounds it is retracted *re* its neutral position. For high sounds, the body of the tongue is raised *re* the /ə/, while for low sounds it is lowered *re* the /ə/. Rounded sounds are produced with protruding lips.

Voicing refers to the presence or absence of vocal fold vibration. The feature 'continuant' distinguishes between stops (nasal stops included) and other sounds. Anterior sounds are produced with the tongue tip at or before the alveolar ridge in the vocal tract, and coronal sounds are produced with the tongue tip or blade raised (it includes some palatal consonants). Strident sounds are fricatives or affricates (=stop + fricative) with high-frequency noise (a hissing sound). Nasal sounds are produced with the velum lowered, resulting in airflow through the oral and nasal cavities.

The notion of distinctive features is not only relevant for analysis purposes but also for the description of phonological processes. Some phonological processes, such as for instance assimilation, affect certain combinations of features more than others,

and this should be reflected in the phonological representation. This is possible if it is assumed that each distinctive feature is free to act independent of the other features it may be associated with. This approach consisting of 'interconnecting levels' (tiers) is known as autosegmental phonology: each feature can influence a neighboring sound segment, irrespective of whether other associated features do so. In particular, feature geometry describes the (nonlinear) action of one feature on another by representing their relationship as hierarchical (tree) structures.

Rules

Phonological rules make statements about which allophones of a phoneme will occur in a specific context. The change of one item to another item in a certain environment can be stated by a rule. The validity of the rule can then be tested against other examples to determine exceptions between sounds within and across languages. Rules are important for discovering the universal principles governing the use of sounds in languages. It is necessary to specify the item(s) affected, the change(s) that takes place (indicated by '→'), and the environment in which the change occurs (indicated by '/'). In linear rule writing, the original state of affairs is given on the left of the arrow and the structural change and environment on the right of the arrow. For example, in English, vowels are nasalized before a nasal stop. This can be stated as follows:

$$\text{e.g. } /æ/ \rightarrow [\tilde{æ}] / \_\_\_ /n/$$

It is also possible to use brace notation to describe one process in two different environments. The following example states that /t/ can be pronounced as a glottal

TABLE 1 Possible Distinctive Features of Some English Vowels (Not Exhaustive)

	u	ɔ	o	ɑ	ʌ	æ	e	ɛ	i	ɪ	ə
Front	−	−	−	−	−	+	+	+	+	+	−
Back	+	+	+	+	−	−	−	−	−	−	−
High	+	−	−	−	−	−	−	−	+	+	−
Low	−	−	−	+	+	+	−	−	−	−	−
Round	+	+	+	−	−	−	+	−	−	−	−

TABLE 2 Possible Distinctive Features of Some English Consonants (Not Exhaustive)

	p	t	k	b	d	g	m	n	f	s	θ	ʃ	v	z	ð	h	l	r	w	j
Voice	−	−	−	+	+	+	+	+	−	−	−	−	+	+	+	−	+	+	+	+
Continuant	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+	+
Strident	−	−	−	−	−	−	−	−	+	+	−	+	+	+	−	−	−	−	−	−
Anterior	+	+	−	+	+	−	+	+	+	+	+	−	+	+	+	−	+	+	−	−
Coronal	−	+	−	−	+	−	−	+	−	+	+	+	−	+	+	−	+	+	−	+
Nasal	−	−	−	−	−	−	+	+	−	−	−	−	−	−	−	−	−	−	−	−

stop before a consonant or the end of a word (such as in 'button').

$$/t/ \rightarrow [\textglotstop]/ \underline{\quad} \begin{Bmatrix} C \\ \# \end{Bmatrix}$$

By using parentheses and braces it is possible to formulate rules of greater complexity. Moreover, the Greek alphabet, or alpha notation, can be used to match features in different places in the rule.

While some rules state how features are affected by the context of other features (feature-changing rules, such as nasalization, glottalization, flapping), other rules can affect the entire segment. For example, in English a schwa is inserted between a final liquid and a nasal, as in [fɪləm] (instead of /fɪlm/). This example of insertion can be described as:

$$\varnothing \rightarrow \text{ə}/ \begin{bmatrix} +\text{cons} \\ +\text{son} \\ -\text{nas} \end{bmatrix} \underline{\quad} \begin{bmatrix} +\text{cons} \\ +\text{nas} \end{bmatrix} \#$$

where 'cons' refers to consonantal sounds, 'son' to sonorant sounds, i.e. sounds that contain considerable energy (such as /m,l,j/), and 'nas' to nasal sounds.

Given the presence of certain features, one can predict the value of other features in that segment. For example, there are no rounded front vowels in English. If a vowel is specified as [-back] it is also [-round]. It is therefore not necessary to specify [-round]. A blank indicates that the feature is predictable by a phonological rule of the language. Similarly, if a phoneme is [+nasal] it is also [+voiced] in English. Sometimes, nasal phonemes are 'devoiced' after an initial /s/ [snʊp]. However, this does not affect the meaning of the word, as is the case in Burmese where /ma/ means 'health' and /mḁ/ means 'order' (therefore, they are different phonemes).

Suprasegmentals

Until now, we have considered the individual phonemes of a sound system. However, several phonological processes affect units that are larger than phonemes, such as syllables, words, phrases, and sentences. In some languages, words contain (with some exceptions) only front or back vowels (e.g. Turkish). The analysis of phonological features in terms of units larger than a phoneme is dealt with by prosodic phonology, and autosegmental phonology. Another branch of phonology, suprasegmental phonology, deals with aspects of pitch, loudness, tempo, rhythm, and tone. A phonological theory concerned with organizing phonemes into groups of relative prominence that emphasizes the relationship between phonemes and rhythm, and intonational stress is known as metrical phonology.

Optimality Theory

Until now, phonological processes have been addressed by means of rules and derivations (derivational Generative Phonology). Application of one rule often affects the subsequent application of some other rule. However, during the last decade, Optimality theory has obtained a dominant role in phonology. This approach has introduced an alternative way of modeling the relationship between words and sounds.

Rather than trying to define how a language's words may be derived by combining given phonemes according to certain rules, Optimality theory approaches the problem from the opposite direction: in principle, any combination of features/sounds is assumed to be possible, but each language imposes certain constraints in a language-specific way on the endless possibilities.

The constraints are assumed to be universal while they may be ranked differently, which allows for language-specific grammars.

While phonology attempts to model knowledge of a language, it should also reflect the fact that languages are learnable. Proponents of Optimality theory argue that apart from dealing with phonological phenomena, it also includes a better learning theory than traditional (derivational) approaches. In addition, Optimality theory has already proven to be a productive new tool in other domains of linguistic analysis, such as morphology, syntax, and even semantics.

References

Anderson, Stephen. 1974. The organization of phonology. New York: Academic Press.

Chomsky, Noam, and Morris Halle. 1968. The sound pattern of English. New York: Harper & Row; Paperback edition 1991, Cambridge, MA: MIT Press.

Clements, George N. 1992. Phonological primes: features or gestures?. Phonetica 49. 181–93.

Durand, Jacques, and Francis Katamba. 1995. Frontier of phonology. London: Longman.

Goldsmith, John. 1990. Autosegmental and metrical phonology. Oxford: Blackwell.

Goldsmith, John. 1995. The handbook of phonological theory. Oxford: Blackwell.

Jakobson, Roman. 1990. On language. Cambridge, MA: Harvard University Press.

Kager, R. 1999. Optimality theory. Cambridge: Cambridge University Press.

Ladefoged, Peter, and Ian Maddieson. 1996. The sounds of the World's languages. Oxford: Blackwell.

Maddieson, Ian. 1984. Patterns of sounds. Cambridge: Cambridge University Press.

Roca, Iggy. 1994. Generative phonology. London: Routledge.

Spencer, Andrew. 1996. Phonology: theory and description. Oxford: Blackwell.

Vihman, Marilyn. 1996. Phonological developments: the origins of language in the child. Oxford: Blackwell.

ASTRID VAN WIERINGEN

Phrase Structure

Sentences are not just linear strings of words. There is ample evidence that words in a sentence can form a unit that excludes other words in the same sentence. Such units are referred to as constituents. Constituents are hierarchically organized, forming larger phrases. For example, a sentence will often consist of a subject and a predicate. The constituent that forms the predicate can in turn consist of several smaller constituents, such as a verb, a direct object, an adverbial, and so on. These constituents, too, may be complex and contain even smaller constituents. The internal structure of sentences and smaller constituents is generally referred to as 'phrase structure'.

Evidence for the assumption that the internal organization of a sentence consists of hierarchically ordered constituents is manifold. One piece of evidence concerns the fact that the words in a sentence can be redistributed in certain ways such that the result is again a grammatical sentence. In such a redistribution process, certain words stay together. This indicates that they form a unit—a constituent. For instance, in the sentence in (1a), the phrase *the films by Pierre Paolo Pasolini* functions as the direct object of the verb *likes*: it is the thing that is liked. It thus forms a semantic unit. That it also forms a syntactic unit is indicated by (1b): the phrase can be placed at the beginning of the sentence as a whole (a process technically known as 'topicalization') and the result is still grammatical. This cannot be done with just any string of words that happen to stand next to each other. This is shown by (1c), where the attempt to topicalize *likes the films*, to the exclusion of the rest of the direct object, results in ungrammaticality. This shows that this string does not form a syntactic unit on which syntactic rules such as topicalization can operate. Thus, *likes the films* is not a constituent of the sentence in (1a). (Ungrammatical examples are marked by an asterisk, following the common convention in the literature).

(1a)   John likes the films by Pierre Paolo Pasolini.
(1b)   The films by Pierre Paolo Pasolini, John likes.
(1c)   *Likes the films, John by Pierre Paolo Pasolini.

Another indication that phrases are structured and consist of constituents is that certain strings of words can be replaced by one word, whereas others cannot. In (1a), *the films by Pierre Paolo Pasolini* can be replaced by the question word *what* (which in English has to be placed at the beginning of the sentence), as

in (2a). Again, this is not possible for *likes the films*, as shown by (2b). Replacement thus gives the same indication of constituency as does redistribution.

(2a)   What does John like? (Answer: The films by P.P. Pasolini).
(2b)   *What does John by Pierre Paolo Pasolini? (Answer: Like the films).

The smallest constituents of a sentence are words, which are further organized into phrases, which in turn are further grouped into larger phrases, and so on.

Phrasal constituents are built around a word that functions as the head of the phrase. The head of a phrase is that word in the phrase which in a sense is its most important part: it cannot be omitted, and it determines most of the syntactic properties of the phrase as a whole. For example, subjects of sentences are very often phrases that are built around a noun. The phrase can consist of just this noun, as in (3a), but more material can be added, as in (3b), where the head *music* is modified by an adjective preceding it, and (3c) where a relative clause following the noun is added. In this way, arbitrarily large phrases can be formed. The head cannot be left out, however, as illustrated in (3d).

(3a)   [Music]$_{NP}$ is his favorite pastime.
(3b)   [Classical music]$_{NP}$ is his favorite pastime.
(3c)   [Classical music written in the twentieth century]$_{NP}$ is his favorite pastime.
(3d)   *[Classical written in the twentieth century]$_{NP}$ is his favorite pastime.

If a phrase is built around a noun, the resulting structure is a *noun phrase*, abbreviated as *NP*. Any phrase headed by a noun has a syntactic distribution that is determined by its head. For instance, phrases headed by a noun may appear in subject, direct object, indirect object positions, etc. In contrast, phrases headed by an adjective (APs) cannot normally function as such.

(4a)   [The girl]$_{NP}$ gave [the boy]$_{NP}$ [a new cd]$_{NP}$.
(4b)   *[Pretty]$_{AP}$ gave the man a new cd.

Adjectival phrases, on the other hand, can act as modifiers to nouns (e.g. the AP *classical* modifies the noun *music* within the NP *classical music* in (3b), whereas NPs do not ordinarily appear in this position (bar exceptional cases like *a London bus*). The distribution of phrases headed by a verb (VPs such as *cook pasta* or *like music*), phrases headed by prepositions

(PPs such as *in the cupboard* and *under the stairs*), and phrases headed by an adverb (AdvPs such as *very quickly* or *yesterday*) is yet different.

Depending on the properties of the head itself, other material within a phrase besides the head can also be obligatory. Certain lexical items require additional elements to be present in a phrase that they are the head of. This can be due, at least in part, to their meaning. For instance, the verb *like* has to combine with an object (5a). Omission of this so-called complement to the verbal head leads to ungrammaticality (5b) in this case. However, not all complements are obligatory. The verb *eat* can combine with an object (5c), but need not necessarily do so (5d). In the latter case, an object is implicitly understood.

(5a) John likes music.
(5b) *John likes.
(5c) John is eating plums.
(5d) John is eating.

Apart from heads and complements, phrases may contain additional, purely optional, material. Such optional elements are referred to as adjuncts. In VPs, they are most often adverbials such as time, place, and manner adverbials; in NPs, modifying relative clauses, for example, can act as adjuncts. Adjuncts differ from heads and complements in not being unique in a phrase. Whereas a phrase can only have one head, and (barring some problematic cases) also only one complement, adjuncts can be freely stacked. This is illustrated in (6).

(6) John often walks his dog in the park, in the spring, on a nice Sunday afternoon, while whistling *The blue Danube*.

An early formalism of expressing the possible constituent structures of sentences made use of so-called rewrite rules (also known as phrase structure rules; see Chomsky 1957, 1965). They have the general form of (7), where XP stands for a phrase of any category, and Y, Z, and W for its constituent parts. The rule in (7) states that the phrase XP consists of Y, Z, and W (in technical parlance, XP is said to be rewritten as Y Z W). Y Z W can be words, but can also be phrases themselves.

(7) XP → Y Z W

Simplifying somewhat, the structure of NPs can be expressed by the rewrite rule (8a). Rule (8a) states that NPs consist of a determiner, optionally an adjective, the head noun, and optionally a PP. The structure of a VP can be given by the rule in (8b), which states that a VP consists of a verb, an NP (functioning as an object), and (optional) adverbs. The structure of a PP can be represented as given by the phrase structure

rule in (8c), which combines a preposition and its complement.

(8a) NP → Det (Adj) N (PP)
(8b) VP → V NP (Adv)
(8c) PP → P NP

It is worth noting that the NP mentioned in (8c) can be rewritten in accordance with (8a), meaning that it can contain a PP. This PP can be rewritten again in accordance with (8c), so that it contains an NP, which can be rewritten yet again in accordance with (8a), and so on, ad infinitum. This results in sequences like in (9). In principle, there is no limit to such sequences, although for practical purposes they will stop at some point. This shows a pervasive property of the phrase structure of natural languages: it is recursive.

(9) I saw the mouse in the hole in the wall of the house next to the river beside the meadow in the county next to …

Sentences are formed by combinations of several phrases, in the simplest of cases an NP that functions as the subject of the sentence and a verb phrase that functions as the predicate (as for instance in [[John]$_{NP}$ [likes Mary]$_{VP}$]$_S$). This is expressed by the rewrite rule in (10).

(10) S → NP VP

Rewrite rules were replaced in favor of a theory of phrase structure, which proved to be very influential. This theory is known as X'-theory (pronounced as X bar theory). It was first proposed in Chomsky (1970), and further elaborated in such works as Jackendoff (1977), Fukui (1986), and Speas (1990). It originated from the idea that all phrases, regardless of the categorial status of their head (i.e. whether they are NPs, VPs, PPs, AdjPs, or AdvPs), are built along the same structural schema, given by the *tree diagram* in Figure 1.

X is a variable standing for a head of any category (N, V, P, A, Adv), YP is the *complement* of the head, while ZP is called its *specifier*. X' and X'' are the *projections* of the head X. When the head X combines with the complement YP, it projects up to the X' (pronounced X bar) level. The X' constituent further combines with a specifier leading to the projection of the X'' (pronounced X double bar) level. In the classical
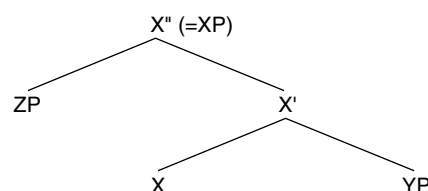
Figure 1

version of the theory, the X" level is considered to be the highest level up to which the head can project. It is therefore also referred to as the XP level, designating a full phrase.

Examples of how phrases of various types fit into the general scheme in Figure 1 are given in Figure 2.

Like the complement and the head, the specifier is unique. For example, while a noun can take a possessive phrase as specifier, and can take a determiner as specifier, it cannot take both:

(11a) Mary's copy of *Dracula*
(11b) a copy of *Dracula*
(11c) *a Mary's / Mary's a copy of *Dracula*

There are several structural relations that can be defined over the tree diagram in Figure 1. Any two constituents that are immediately dominated by the same node are called sisters. X and YP are sisters, being immediately dominated by the X' node, as are X' and ZP, the latter two being immediately dominated by the XP node. The XP node is said to be the mother of X' and ZP, while the X' node is similarly the mother of the X and YP nodes. The specifier of the head X can therefore be defined as the immediate
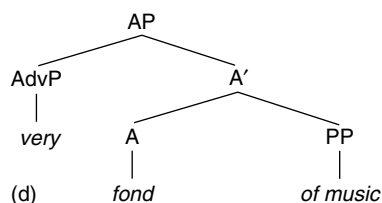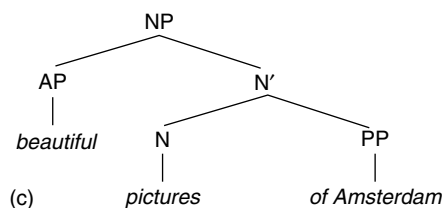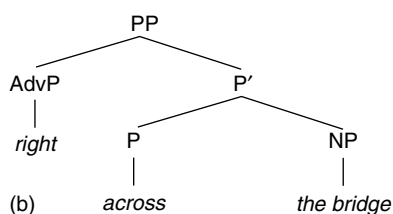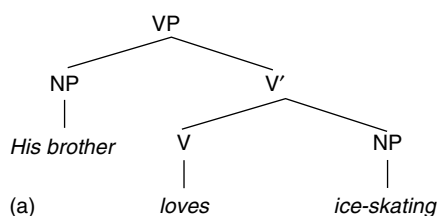
daughter of X" and sister of X', and the complement can be defined as the immediate daughter of X' and sister of X.

The schema in Figure 1 cannot accommodate all the possible constituent parts of phrases. In particular, it cannot accommodate the adjuncts in a phrase, of which there can be arbitrarily many (see above). To accommodate adjuncts, it is assumed that bar levels can be reiterated. When an adjunct is added to a phrase, the head does not project to a higher bar level; instead, the bar level is just repeated. In this way, an adjunct that is added at the double-bar level can be formally distinguished from the (unique) specifier of the head: whereas both the adjunct and the specifier are daughters of an X"-node, only the specifier is the sister of an X'-node; the adjunct is the sister of an X"-node (Figure 3).

Because the specifier can also be distinguished from adjuncts by the very fact that adjuncts can be stacked, whereas the specifier is unique, it is not clear that it is essential to distinguish between the two formally as well (see below).

An assumption built into most modern phrase structure theories is that all nodes are maximally binary branching. This means that no node can have more than two daughters. If this assumption is adhered to, 'flat' structures of the type in Figure 4, for a VP that contains both a direct object and an indirect object, are impossible.

The consequence is that there must be a more hierarchical structure inside VP. Evidence for this is said to come from examples like (12a, b). These show that the indirect object can act as antecedent for a reflexive element (for instance, *himself*) inside the direct object,
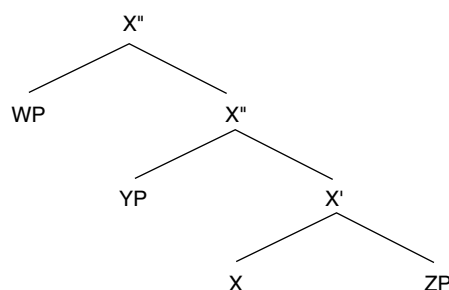
Figure 2
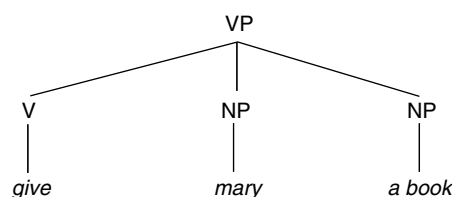
Figure 3

Figure 4

but not the other way around. Evidence from other types of sentences shows that the antecedent for the reflexive has to be higher in the structure than the reflexive itself. The data in (12a, b) therefore indicate that the indirect object and the direct object are not on the same level in the structure. Instead, the indirect object must be higher than the direct object.

(12a) Mary showed Bill a double of himself.
(12b) *Mary showed a double of himself Bill.

The scheme in Figure 1 is not intended to express a linear order of constituents. The theory of phrase structure in the first instance is about the hierarchical relationship between phrases. Different principles must then determine whether a complement precedes or follows the head, as well as whether the specifier precedes or follows the head.

Syntactic theory distinguishes between two classes of words: lexical and function words. The position of function words in phrase structure theory has always been rather unclear. Recently, such elements have come to be regarded by some as heads of phrases, on par with lexical categories. In line with this development, items like determiners (such as *the* and *a* in English), complementizers (such as *that*, *if*, *whether*), auxiliary verbs and modals (such as *to be*, *can*, *will*), and even inflectional affixes for tense and agreement (such as the English *–ed* suffix in past tense forms) are regarded as the heads of phrases. This view has various consequences for how the internal structure of phrases is supposed to be built up. For instance, a phrase like *the house* is no longer seen as an NP, but rather as a determiner phrase (DP), headed by the determiner D, which takes the NP *house* as its complement, as represented in Figure 5.

Similarly, sentences are now regarded as projections of the inflection we see on the verb, and hence as inflectional phrases or IPs (Figure 6). Subjects are in the specifier position of IP. In the same vein, sentences introduced by a complementizer are regarded as complementizer phrases or CPs. The complementizer head in the CP takes an IP as its complement, which in turn takes a VP as its complement (Figure 7). The specifier position of CP can also host elements in some sentences, such as question words.

The proliferation of functional categories in the theoretical inventory also has consequences for the notion of adjunction. With every additional functional projection, an additional specifier position becomes available. This means that it is at least possible to discard the assumption that bar levels have to be repeated in order to accommodate adjuncts. Instead, every adjunct may be regarded as a specifier of a particular functional projection, as extensively argued in Cinque (1999).
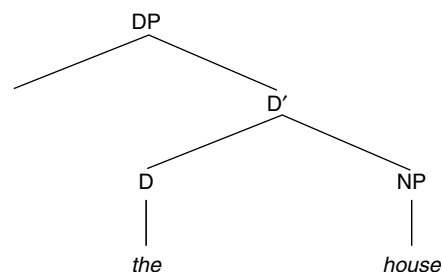
Figure 5
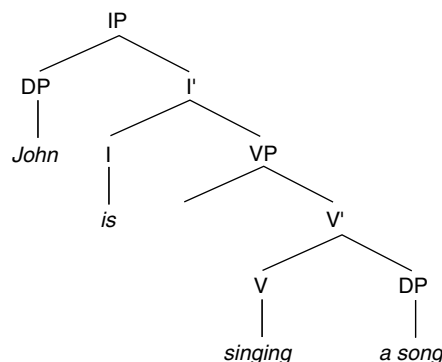
Figure 6
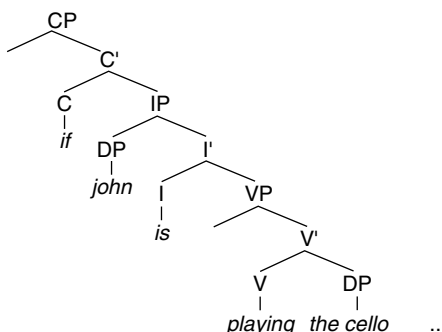
Figure 7

Classical X'-theory not only needs the possibility of repeating bar levels to accommodate all the possible constituents of a phrase, but it suffers from the opposite problem as well: it sometimes forces more syntactic positions to be present in the structure than are actually necessary. For instance, a head can take a specifier while not taking a complement. Whereas (13b) is as well formed as (13a) is, in classical X'-theory there nevertheless is a complement position present in (13b) as well.

(13a) John's collection of mushrooms
(13b) John's collection

It is even possible that a phrase contains just its head, without a specifier or a complement, as for instance in (14b) (vs. (14a), with specifier and complement present). If phrases always have the form of

Figure 1, then in such cases there is both a specifier position and a complement position, neither of them containing any lexical material.

(14a) (I expect) [Mary to win the race]
(14b) (I expect) [to win]

Partly because of this inelegant aspect of X'-theory, a more flexible view of phrase structure has been adopted (see Speas 1986; Chomsky 1994), usually referred to as Bare Phrase Structure theory. Its basic assumption is that a head projects as often as it combines with another phrase, no more and no less. One consequence of this is that it is no longer possible to distinguish between specifiers and adjuncts in structural terms (see also the previous point on functional structure). Instead, the specifier can be distinguished relationally, since it establishes a unique relation with a head.

References

Abney, Stephen. 1987. The English noun phrase in its sentential aspect. Ph.D. Dissertation. Cambridge, MA: MIT.

Cinque, Guglielmo. 1999. Adverbs and functional categories: a cross-linguistic perspective. New York, Oxford: Oxford University Press.

Chomsky, Noam. 1957. Syntactic structures. The Hague: Mouton de Gruyter.

———. 1965. Aspects of the theory of syntax. Cambridge, MA: MIT Press.

———. 1970. Remarks on nominalization. Readings in English transformational grammar, ed. by Roderick A. Jacobs and Peter S. Rosenbaum. Waltham, MA: Ginn.

———. 1994. Bare phrase structure. Cambridge, MA: Distributed by MIT Working Papers in Linguistics.

Fukui, Naoki. 1986. Theory of projection in syntax. Stanford, CA: CSLI Publications, Center for the Study of Language and Information, and Tokyo: Kurosio Publishers.

Grimshaw, Jane. 2003. Words and structure. Chicago: University of Chicago Press.

Jackendoff, Ray. 1977. X' Syntax: a study of phrase structure. Cambridge, MA: MIT Press.

Kayne, Richard S. 1984. Connectedness and binary branching. Dordrecht: Foris.

Pollock, Jean-Yves. 1989. Verb movement, universal grammar, and the structure of IP. Linguistic Inquiry 20. 365–424.

Speas, Margaret. 1986. Adjunctions and projections in syntax. Ph.D. Dissertation. Cambridge, MA: MIT.

AMELA CAMDZIC AND PETER ACKEMA

Pidgins and Creoles

Although until recently creolists were more or less agreed on the definition of 'pidgin' and 'creole', this is no longer so. Also, the adage that 'a creole is a nativized pidgin' (Hall 1966) is no longer universally accepted. It has been observed, for example, that at least some of the French-lexicon creoles do not have a pidgin ancestry (Chaudenson 2001). Further, the more we learn about the transition from pidgin to creole, the less certain we are about where to draw the line between the two (Baker 1995). At the same time, creolists are becoming aware that it may not be justified to group pidgins and creoles together, even though in the past they have often been treated as if they belong to one category. In spite of all these difficulties, some definitions will be presented here, if only to give the reader an idea of the distinctions. A pidgin is a nonprimary language that is the result of language contact. A creole is a primary language that is the result of language contact.

Before explaining these definitions, a brief survey of the pidgins and creoles of the world is useful. According to the most extensive survey that has been made until now (Smith 1995), the total number of pidgins and creoles is around 350, including both the ones that are extinct as well as varieties that are intermediate between a 'true' creole and its related 'lexical donor language' (also called 'lexifier language', 'superstrate language', or simply 'superstrate'). The total number of speakers, including those who only use a creole as a second language (a pidgin is a second language by definition), may be estimated at around 100 million. By far, the largest group, especially among the creoles, is formed by those that have derived the bulk of their lexicon from an Indo-European language such as Portuguese, French, English, Spanish, and Dutch. It is no coincidence, of course, that these are the languages spoken by the European nations that played the leading role in the European expansion. It was the contact between the European languages of the explorers and colonizers, on the one hand, and the non-European languages of the people with whom they came into contact, on the other (referred to as 'substrate language(s)', or simply 'substrate'), that gave rise to the emergence of these pidgins and creoles. Other Indo-European languages, apart from the ones mentioned above, that have played a role in the emergence of pidgins and creoles include German, Russian, Italian, and Hindustani.

Although these European-lexicon pidgins and creoles have attracted the lion's share of the attention of

linguists, there is a significant and linguistically just as important group of pidgins and creoles whose lexicons are based on non-Indo-European languages. The lexifier languages in this group include Arabic, Japanese, Chinese, and a number of Bantu, Austronesian, Australian, Papuan, and Amerindian languages. An important difference with the group of European-lexicon pidgins and creoles is that in many cases no Indo-European language was involved in their formation. Because of this, the study of non-European pidgins and creoles may serve to correct any biases that may have resulted from the focus on European pidgins and creoles that has characterized the field of pidgin and creole linguistics for a long time (see Thomason 1997, 2001). An important question in this regard is whether the features of pidgins and creoles that are sometimes regarded as universal might not be an artifact, resulting from the fact that all the major lexifier languages are structurally similar, belonging as they do to only two branches (Germanic, Romance) within one and the same language family, Indo-European. To get an idea of the variety as well as the geographical distribution of the world's pidgins and creoles, the maps in Holm (2000) are a good starting point. As far as the Pacific is concerned, much more detailed information can be gleaned from the splendid three-volume language atlas edited by Wurm et al. (1996). Unfortunately, a similar work for the Atlantic region is not (yet) available.

As mentioned earlier, the definitions given above require some comments. First of all, what is meant by 'primary' and 'nonprimary' language? If you are multilingual, your primary language is the one with which you feel most at home, the one you use most often, even though it may not be your native language. When we say that a pidgin is a nonprimary language, it means that pidgin-speakers have at least one other language in their repertoire—their primary language— and that the pidgin is used only as an auxiliary language, when communicating with speakers with whom they have no other language in common. Creoles, on the other hand, are primary languages by definition, although, again, this does not mean that they are necessarily native languages. Second, if pidgins are nonprimary languages, is it justified to view them as 'true' languages in the first place? Although this is an interesting question in itself (when is something a language?), it is too complex to be dealt with here. For the purpose of this essay, however, pidgins are referred to as languages. As for creoles, there is no question about it: they are languages just like any other.

The one thing that pidgins and creoles have in common, according to the definitions given above, is that they are a product of language contact. Language contact, therefore, is the key concept in these definitions: it is a necessary, although not sufficient, condition for pidginization and creolization (the process by which a

pidgin or a creole comes into being). This means, first, that there will be no pidginization or creolization without contact, and, second, that contact does not automatically lead to pidginization or creolization. When we look at pidginization and creolization as language contact phenomena, the following questions present themselves:

(1) What other types of language contact are there, apart from pidginization and creolization?
(2) In what ways do pidginization and creolization differ from other types of language contact?
(3) What are the factors that decide whether language contact will lead to pidginization or creolization?

These three questions will guide the remainder of the discussion.

What other types of language contact are there, apart from pidginization and creolization?

The linguistic effects of language contact are dependent on a number of factors, the most important of which is whether or not imperfect language learning plays a role in the contact situation. On the basis of this, two types of language contact can be distinguished (Thomason 2001):

(a) *Shift-induced interference*. When people give up their native language while adopting another one, the new language often displays features from their native language. In this case, the changes in the affected language are a result of imperfect learning.
(b) *Borrowing*. When people maintain their native language, it may still be influenced by another language. This happens when native speakers incorporate features from another language into their native language. In this case, the changes in the affected language are not a result of imperfect learning.

The crucial difference between the two is that in the case of shift-induced interference, speakers introduce changes in a language that is not their native language, while in the case of borrowing they introduce changes in a language that *is* their native language. Although the end-results of these two types of language contact are sometimes hard to distinguish, the processes by which these results come about are very different, for example, in the order in which the different components (lexicon, phonology, etc) of the language system are affected. While borrowing always starts with words and only involves the grammatical system later, it is the other way around in shift-induced interference. The results of the two processes may range from minor changes in the lexicon, such as adoption of a loanword, to major structural changes, such as changes in basic word order.

In what ways do pidginization and creolization differ from other language contact phenomena?

Pidginization and creolization differ from other language contact phenomena (except, perhaps, so-called 'mixed languages'; see relevant lemma) in that neither pidgins nor creoles can be justifiably viewed as changed versions of the languages in contact, while this *is* the case with borrowing and shift-induced interference. In other words, pidgins and creoles are autonomous languages. This entails that they have a grammar of their own, one that is not derivative of or dependent on a pre-existing language. The autonomy of the grammar is the result of a process whereby grammatical 'components' (rules, patterns, processes) that were initially related to one or more of the contributing languages, have become independent. The transition from dependency to independency occurs when these components start interacting with one another to produce a new grammar, rather than forming an incoherent collection of elements derived from other languages.

The idea of pidgins and, especially, creoles as autonomous languages is shared by most creolists, including those who view creolization in terms of either borrowing or shift-induced interference. The view of creolization as a case of (extreme) borrowing is held by Lefebvre (1998), who sees Haitian Creole as the result of a process of relexification, whereby words from the slaves' native language(s) were replaced by similar words from French, while the grammatical structure of the native language(s) was maintained. A very different view is represented by Chaudenson (2001), who believes that—especially the French-lexicon—creoles are the result of a process of language shift (from the slaves' native languages to French), whereby some features may have been introduced as a result of interference. No matter how far apart, both views agree that creoles are autonomous languages, not modified versions of one or more of the contributing languages. The idea of creoles as autonomous languages is also inherent in Thomason's (2001) view of pidginization and creolization as cases of 'abnormal' language transmission. While in 'normal' transmission both the lexicon and the grammatical apparatus are transmitted from one and the same 'parent language' to one or more 'daughter languages,' in cases of 'abnormal' transmission, there are at least two 'parents', one for the lexicon and one for the grammatical system. As a result of this, a pidgin or a creole cannot be a variety of either one of the parent languages; therefore, it is an autonomous system. Finally, in generative approaches to pidginization and creolization (e.g. DeGraff 1999), the question of autonomy hardly arises, as every language—whether pidgin, creole, or something else—is seen as an instantiation of Universal Grammar. Although this brief discussion does not exhaust the list of theories of creole genesis,

it seems fair to say that they all share the concept of autonomicity as a basic feature of pidgins and creoles.

What are the factors that decide whether language contact will lead to pidginization or creolization?

This is one of the fundamental questions regarding pidginization and creolization. Since little attention has been devoted to it, especially from a historical point of view, the discussion here is necessarily tentative. It should also be noted that what follows is restricted to European-lexicon pidgins and creoles (by far the largest group), and that it is not necessarily valid for pidgins and creoles that are lexically based on other languages. It is hoped, however, that insights into the conditions for the genesis of the European-lexicon pidgins and creoles will also shed some light on the other cases.

The European pidgins and creoles both arose in the historical context of the European expansion. This particular phase in history began around 1430 when Prince Henry ('the Navigator') started sending out Portuguese ships to explore the northwestern coast of Africa. Although the exact date when the contacts between Europeans and Africans for the first time led to the emergence of a pidgin or creole language in not known, it is clear that the genesis of the European-lexicon pidgins and creoles took place in the context of the European expansion, whether in Africa, America, Asia, or the Pacific. The age of European expansion was a very special, perhaps even unique, phase in the history of the world, which may be characterized by the following features:

(a) The introduction of trade and production on a truly global scale.
(b) The presence of contact situations characterized by unusually wide social, cultural, and psychological gaps between the parties involved (Europeans on the one hand and Africans, Asians, Native Americans, and Polynesians on the other).
(c) The use of forced labor and forced displacement of large numbers of people (slave trade, indentured labor).
(d) The creation of a new type of society known as 'the plantation complex' (Curtin 1998).

With regard to the last feature, it should be added that, although plantations in the strict sense were not present in every single situation where pidgins and creoles developed, the concept of 'plantation complex' may be interpreted in a sufficiently wide sense to encompass the latter situations as well.

Although the exact relationship between the presence of particular external conditions on the one hand and the emergence of pidgins and creoles on the other is not very clear, a few things may still be said about it. For example, there is the issue of access to the target language. With the increasing disproportion between the

European and non-European segments of the population and the concomitant change from the small-scale 'homestead society' to the large-scale 'plantation society' (Chaudenson 2001), access of the non-Europeans to the European language decreased. However, this cannot be the whole story because there are a number of cases, especially in Spanish colonies, where such disproportion did not lead to pidginization or creolization (McWhorter 2000). Other factors, such as lack of motivation on the part of the enslaved to learn the language of their masters (Baker 1990), may have been involved as well. However, there is another, more elusive, factor that has received little attention until now. If we look at the period of the European expansion from a contemporaneous rather than from a modern perspective, we cannot help but recognize that this unique phase in history confronted all those who were involved in it with an entirely 'new world'. Never before had so many different parts of the world been in contact on such a large and intensive scale. Although the fact that this new contact situation was accompanied by a true 'explosion' of new languages does not necessarily mean that the one was caused by the other, the synchronicity of the two 'events' is striking. It is hoped that future research, especially into the historical dimension of pidginization and creolization, will shed more light on this issue.

References

Arends, Jacques, Pieter Muysken, and Norval Smith (eds.) 1995. Pidgins and creoles: an introduction. Amsterdam, Philadelphia, PA: Benjamins.

Baker, Philip. 1990. "Off target?" Journal of Pidgin and Creole Languages 5. 107–19

———. 1995. Some developmental inferences from historical studies of pidgins and creoles. The early stages of creolization, ed. by Jacques Arend, 1–24. Amsterdam, Philadelphia, PA: Benjamins.

Chaudenson, Robert. 2001. Creolization of language and culture. Revised in collaboration with Salikoko Mufwene [translation of Des îles, des hommes, des langues. Paris: L'Harmattan. 1992]. London and New York: Routledge.

Curtin, Philip. 1998. The rise and fall of the plantation complex, 2nd edition. Cambridge: Cambridge University Press.

DeGraff, Michel, (ed.) 1999. Language creation and language change: Creolization, diachrony, and development. Cambridge MA: MIT Press.

Hall, Robert. 1966. Pidgin and creole languages. Ithaca, NY: Cornell University Press.

Holm, John. 2000. An introduction to pidgins and creoles. Cambridge: Cambridge University Press.

Lefebvre, Claire. 1998. Creole genesis and the acquisition of grammar: the case of Haitian Creole. Cambridge: Cambridge University Press.

McWhorter, John. 2000. The missing Spanish creoles: Recovering the birth of plantation contact languages. Berkeley CA: University of California Press.

Smith, Norval. 1995. An annotated list of creoles, pidgins, and mixed languages. ed. by Jacques Arends, Pieter Muysken, and Norval Smith, Pidgins and creoles: an introduction, 331–74. Amsterdam, and Philadelphia, PA: Benjamins.

Thomason 1997. Thomason, Sarah, (ed.). Contact languages: A wider perspective. Amsterdam/Philadelphia PA: Benjamins.

Thomason, Sarah. 2001. Language contact: an introduction. Philadelphia: Georgetown University Press.

Wurm, Stephen, Darrell Tryon, and Peter Mühlhäusler (eds.) 1996. Atlas of languages of intercultural communication in the Pacific, Asia and the Americas, 3 vols. Berlin: Mouton de Gruyter.

JACQUES ARENDS

See also **Language: Contact-overview; Lexical Borrowing**

Pike, Kenneth Lee

Kenneth Pike, like his predecessors, based his theoretical contribution firmly on linguistic field work. He followed their lead in attempting to explain the ways in which the languages of North America structured words, phrases, and clauses. Unlike his teachers, he added a needed emphasis on meaning and context to earlier analytical methods. Pike's interest in linguistics began when he attended the second session of linguistics training led by William Cameron Townsend in 1935 at a remote Arkansas farm. After this rudimentary training, Pike applied his new knowledge to the study of the complex tone patterns, morphology, and syntax of the Mixtec language in the highlands of Oaxaca, Mexico. The following year (1936), he taught the phonetics component of the Townsend's course, using his personal field experience to explain phonetic concepts. His introduction to the initial writings of Edward Sapir (1921) and Bloomfield (1933) motivated him to attend the 1937 Linguistics Institute of the Linguistics Society of America, where he also met the leading American linguists of the day. He studied under Sapir from 1937 to 1942 at the University of Michigan, focusing on phonetics and descriptive linguistics. Pike was later hired by the university and taught linguistics there from 1948 to 1979.

Pike spent much of his early efforts (1935–1948) on phonetics and phonology (phonemics) and published textbooks in these areas that were used extensively for many years. His personal field studies and his experience as a consultant to other researchers in Latin America led Pike to investigate phonetics and phonology (then called phonemics). His *Phonetics* (1943) was the most complete book of its kind. Pike's phonetic studies provided insights into what he called contrast, variation, and distribution. Contrast refers to the ways in which sounds are different from one another in a particular language. Variation describes the fact that individual sounds are not always pronounced the same way, although native speakers perceive them as the same sound. Distribution deals with the locations in which a sound appears. These three themes continued throughout Pike's writing and formed the basis of his later work on morphology and syntax as well.

Pike pioneered the study of tone languages and helped many other linguists analyze the diverse tone systems of the world. One important contribution was using a constant tone pattern or frame to analyze similar tones. By keeping the frame (such as 'This is a…') constant, linguists could better hear the contrastive tones of the words being studied.

Pike's studies in intonation were likewise an essential tool for many scholars and people involved in teaching English to speakers of other languages. Pike applied intonation studies to poetry as well and published examples of poems (his own included) marked for intonation patterns reflecting how the poem was read by different speakers.

In 1936, Pike developed a 'monolingual demonstration' in which he showed students how to learn a language without relying on English translation. Using a variety of sticks, leaves, and other common objects, Pike would elicit language data from a speaker of a foreign language without using an intermediary language. Within an hour, Pike would be able to explain major features of the sound system, morphology, and syntax of the language. The lecture was very popular, both on campus and in nonacademic settings. It provided a clear illustration of the role of a hypothesis and experimentation in linguistic studies. In 1977, the University of Michigan distributed a video version of one such demonstration of Pike working with a Javanese speaker.

Pike is best known for his work in tagmemics, a theoretical approach to morphology and syntax that integrates language forms and grammatical functions. Based on the concepts of sounds (phonetics) and sound patterns (phonemics), Pike proposed that there were similar features in morphology (morphemes) and syntax (syntagmemes). He defined the tagmeme as a combination of particular grammatical 'slots' or positions with the elements that could 'fill' that position. In English, for example, the subject position (slot) of a sentence can have (be filled by) a Noun Phrase, a pronoun, or proper noun, but not by an adverb or a conjunction. Eventually, Pike would add semantic information to his tagmemic model. In the previous example, Pike noted that the role or function of a Subject in English can be either the person who does the action (an Agent) or the person who is affected by the action (a Patient or Undergoer), as in a passive clause.

Using principles of contrast, variation, and distribution from phonemics (phonology) and phonetics, Pike coined the terms 'emic' and 'etic' in 1954 to discuss word and clause structures. Pike's seminal work *Language in relation to a unified theory of the structure of human behavior* presents his philosophy of language and culture that formed the basis of the emic/etic distinction later used in anthropology and other fields. With his wife and coteacher, Evelyn, Pike developed a unified approach for studies of morphology, syntax, and discourse at a time when other linguists limited syntax to the clause or sentence level. Their *Grammatical analysis* (1982) has been widely used as a textbook in descriptive linguistic methodology.

Pike's early students at the University of Michigan include Alton Becker, Ruth Brend, Charles Fillmore, John Gumperz, Ilse Lehiste, and Velma Pickett. His pioneering work in tagmemics has also been developed by Walter Cook, Soedarjanto, Linda K. Jones, and Robert Longacre, among others. Other linguistic theories overshadowed Pike's work at times, but his teaching and writing influenced hundreds of field linguists and helped them to document indigenous languages around the world. After his retirement from the University of Michigan in 1979, Pike served as an adjunct professor of linguistics at the University of Texas–Arlington.

Pike's contribution to linguistics was overshadowed when tranformational approaches to syntax became popular. Although tagmemics handled a larger portion of the overall linguistic analysis, it did not handle such areas as active vs. passive transformations or the ways in which certain patterns can be derived from other, more basic, patterns.

Pike was one of the first people involved in the Summer Institute of Linguistics (SIL), a Christian organization that sponsors linguistic and cultural research as well as literacy and translation programs in indigenous language communities. When Pike was not teaching at the University of Michigan, he split his time between his long-term research among the Mixtec people and helping his SIL colleagues. In 1951, Pike and his coworkers published a translation of the New Testament in San Miguel Mixtec, the first such translation completed by SIL. From 1936 through 2000, he was regularly involved with SIL schools, first in

Townsend's training program in Arkansas, then at the University of Oklahoma (1941–1987), and later at other locations. As President of the Summer Institute of Linguistics from its incorporation in 1942 until 1978, Pike oversaw the development of SIL's worldwide linguistic research and training in over fifty countries. Pike personally supervised the beginnings of the SIL schools at the University of Oklahoma (1941), in Australia (1950), and in Great Britain (1951). He was also active in the Linguistic Society of America and the Linguistic Association of Canada and the United States.

Pike was the author of more than 20 books and 200 articles, including articles on the interaction of his personal spiritual beliefs and his academic work. He was also a prolific poet and frequently incorporated his poems in his linguistic writing and teaching. He was an outstanding consultant who assisted other linguists to see the patterns in the languages they studied. He frequently coauthored articles with younger colleagues and was coeditor of linguistic volumes in such diverse settings as Indonesia, Nepal, and Papua New Guinea. Pike lectured in 43 countries covering five continents and received eight honorary degrees and numerous other honors including the Presidential Merit Medal from the Philippines (1974). His work with SIL led to his being nominated for the Nobel Peace Prize each year from 1982 to 1996, based on the positive impact of the linguistic, literacy, and translation efforts of SIL in over 1,000 languages.

Biography

Kenneth Lee Pike was born in Woodstock, Connecticut on June 9, 1912. He received his Th.B. (1933) and Ph.D. (1942) for studies in phonetics, mentored by Edward Sapir, University of Michigan. He was a member of the Summer Institute of Linguistics, 1935–2000; President, 1942–1979; and President Emeritus, 1979–2000. He was also Associate Professor of Linguistics, University of Michigan, 1948–1953; Professor of Linguistics, 1954–1979; Chair of Linguistics Department, 1974–1979; and Professor Emeritus 1979–2000. He was Permanent Council Member,

International Phonetic Association, 1945; member, Linguistics Society of America; and President, 1961. Pike was a member of the Linguistics Association of Canada and the United States and its President in 1977. He was also a member, American Academy of Arts and Sciences, 1973 and member of the National Academy of Science, 1985. He was nominated for the Nobel Peace Prize in 1982–1996. He received Fulbright lectureship in the USSR in 1988. Kenneth Lee Pike died in Dallas, Texas on December 31, 2000.

References

Brend, Ruth M. (ed.) 1972. Selected writings to commemorate the 60th birthday of Kenneth Lee Pike. The Hague: Mouton.

Brend, Ruth M. (compiler) 1987. Kenneth Lee Pike bibliography. Bloomington, Indiana: Eurasian Linguistic Associa-tion.

Brend, Ruth M., and Kenneth L. Pike (eds.) 1977. The Summer Institute of Linguistics: its works and contributions. The Hague: Mouton.

Headland, Paul, Kenneth L. Pike, and Marvin Harris (eds.) 1990. Emics and etics: the insider/outsider debate. Newbury Park, CA: Sage.

Pike, Kenneth Lee. 1943. Phonetics: a critical analysis of phonetic theory. Ann Arbor, MI: University of Michigan Press.

———. 1945. The intonation of American English. Ann Arbor, MI: University of Michigan Press.

———. 1947. Phonemics: a technique for reducing languages to writing. Ann Arbor, MI: University of Michigan Press.

———. 1948. Tone languages. Ann Arbor, MI: University of Michigan Press.

———. 1967. Language in relation to a unified theory of the structure of human behavior. The Hague: Mouton.

———. 1982. Linguistic concepts. Lincoln, NB: University of Nebraska Press.

Pike, K.L., and Evelyn Pike. 1982. Grammatical analysis. Dallas, TX: Summer Institute of Linguistics.

———. 1993. Talk, thought, and thing: the epic road to conscious knowledge. Dallas, TX: Summer Institute of Linguistics.

———. 1977. Seasons of life: a complete collection of Kenneth L. Pike's poetry, compiled and ed. by Sharon Heimbach. Huntington Beach, CA: Summer Institute of Linguistics.

Pike, Eunice V. 1981. Ken Pike: Scholar and Christian. Dallas, TX: Summer Institute of Linguistics.

The Summer Institute of Linguistics provides additional information about Kenneth Pike at http://www.sil.org/klp/.

PETER J. SILZER

See also **Mexico; Tone Languages**

Pintupi and Pama-Nyungan Languages

The Pintupi language is an Australian indigenous language spoken in the Western Desert, eastern Central Australia, which is situated approximately 400 km northwest of Alice Springs in the Northern Territory. It is mainly spoken in and around the Papunya Indigenous Community Settlement. Pintupi belongs to a group of

languages which includes Ngaanyatjarra, Pitjantjatjara/ Yankunytjatjara, Luritja/Pintupi, with the last two language groups being very similar in formation (Simpson 1993:136, 142). These languages form part of the Pama-Nyungan grammar group of Australian languages and exclude other indigenous languages from the north of Western Australia and the north and northwest of the Northern Territory (Yallop, C., p. 16). According to an Inquiry by the Australian Government in 1992, 90 of the original 250 Australian indigenous languages are still living languages and approximately 3,000+ people still speak the eastern Western Desert group of languages. The Western Desert, western Central Australia group of languages include Manjiljarra, Yulparija, Kukatja, Ngaanyatjara, and Ngaatjatjarra languages, with approximately 1,000+ people speaking these languages. This brings the total number of people speaking the Western Desert languages of Central Australia to approximately 5,000, making it the largest group of speakers of traditional indigenous languages in their natural environment in Australia. In 1995, Peter Trudgill wrote in his book *Sociolinguistics*:

> In Australia, for instance, there used to be about 200 aboriginal languages. Of these, 50 are already dead, and another 100 are very close to extinction. Perhaps as few as 30 will make it to the year 2000. (p. 177)

It has therefore become an important task for linguists today to do as much as they can to work toward the preservation of these endangered languages. Australian indigenous people have been either discouraged or forbidden to speak their own languages since colonization, and many Aboriginal languages have been lost to the world. These languages are 'oral' only and have no written form apart from drawings, such as is found in caves, in their natural settings, and it is only in recent times that linguists have been able to do more extensive fieldwork to give written form to the surviving few indigenous languages. Pintupi is one of these languages, and translators, such as Aboriginal Bible translator Ken Hansen, who lived in a Pintupi community for several years, have done significant work in this area.

In his article on 'Translating for the Pintupi', Ken Hansen says that when he first began to learn to speak Pintupi his progress was very slow because the sounds of the indigenous language were more complex than the sounds of English. It was therefore very difficult to transcribe into a written form.

For example, Hansen says the Pintupi language has words such as:

(1)   ngarrinpa   lying
        and ngarinpa   standing
(2)   wangka   talk
        and wanka   alive

While these words seem very similar in structure, just the addition of a longer 'r' sound gives a completely different meaning to the word. The sound 'ng' is also often used at the beginning, or in the middle, of words, making the language difficult to pronounce for English speakers because we are used to using this sound at the end of words, such as in the word 'sing' (Walsh and Yallop 1993:xi).

Hansen also found that it was very important to use the correct words in different social situations, and because of the strange combination of sounds, this was difficult for him to do. His goals were based on the use of the Bible in traditional indigenous language and culture, and he was obviously very committed to his work. He had estimated that it would take him at least 15 years to work out how to write the language by just listening and transcribing the sounds into an alphabet, a dictionary, and a grammar, let alone have enough information about the culture to translate the Bible. There is now a Pintupi/Luritja to English Dictionary available that has been written by him.

Another enormous problem in doing this type of translation is that the Aboriginal people had never heard of things such as a synagogue, and similar concepts such as 'meeting place' had to be used instead. The indigenous word for 'ceremonial offering', 'kunatinpa', could possibly have been used or incorporated here.

Some other examples of Hansen's Pintupi/English translations are (with literal translations and/or explanations):

synagogue	tjuwuku tjaatji
	the Jew's church
king	mayutju pulka
	big boss
angel	nganka ngurrara
	one who belongs in the sky
pharisee	tjuwuku luwuku mikunytju
	a lover of the Jew's law
Holy Spirit	katutjaka kurrunpa
	God's personal spirit
God	katutja
	the one who pertains above
Christ	kirritja
	Christ pronounced in the Pintupi way
Christ	katutjalu tjamatatjunkula wantirriyantja
	the one whom God promised to send long ago

As can be seen from these translations, the effect of British colonization has been that many words that originally come from English origins would not have been in existence in indigenous language prior to this event. They have since been added to the Australian indigenous language and culture. This is particularly

the case with words such as 'big boss', because many indigenous men became stockmen on cattle properties, and the owner was called the 'big boss', 'mayutja pulka', which was then transcribed into Traditional indigenous language. This is a concept that is important to modern social use of indigenous language and has meant the creation of new Aboriginal word usage (Simpson 1993:137–78). One of these words is the Western Desert word for 'spirit', 'kardiya', which is now also used to refer to 'white people'.

As can be seen in modern word usage in indigenous language is the current, and extremely significant, social event taking place in Australia today, the Reconciliation Process, and the use of the word 'sorry'. The Indigenous people are asking the Australian Government and the Australian people as a whole to say 'sorry' for the past injustices that have been done to the Australian Indigenous people (Haebich 2000:569). These injustices include taking the indigenous children from their natural parents and communities and placing them with white families 'for their own good' (Haebich 2000:567–8). Although the present Australian multicultural population was not specifically involved in perpetrating these injustices, it is seen as a token of respect toward the indigenous people, and an acknowledgement of the shame that the people who were responsible for these injustices should feel, for committing these acts. The associated feelings for this can be incorporated into words such as 'kunta', or 'shame', by saying 'sorry' to what is now left of the indigenous community. Part of the meaning of the Pintupi word 'kunta' relates to the English words and concepts for 'shame' and 'respect'. When used to mean 'respect', it is usually symbolic of 'shyness' in not wanting to show disrespect, such as refusing a person to his face without excuses. Someone who is being respectful of others shows embarrassment by saying whatever they have is less significant than the other persons' possessions, and this type of shame is called 'kuntarrinpana tjanampa'. Children are socialized by shaming, and a child is said to be deaf or unheeding, 'patjarru', when young, but needs to be respectful when older and to be aware of 'kunta' as they learn of the need to show respect within the community (Mawr, B:126–157). These emotional concepts have a great significance in the Pintupi understanding of social behavior and it could be further said that this concept is of importance to the indigenous people as a whole. It can then be an integral part of the Reconciliation Process of saying 'sorry', therefore incorporating an important indigenous concept of being sorry for the bad things that someone has done within the indigenous concept of their own culture.

The following are Pintupi concepts with their English meanings relating to feelings:

kunta	shame
rama	mad
ngulu	fear
rarru	anger
pukulpa	happy
ngarru	happy

These concepts have been interpreted by researchers, such as Meyers, as:

(1) 'Mad, crazy persons are said to be "deaf", "rama", because they take no notice of what is being said to them. "Rama", or "not hearing", therefore relates to people who do not take notice of advice and are therefore not thinking.'

(2) 'The English and Pintupi concepts of "fear" or "ngulu" are the same and generally mean to be frightened or afraid of the consequences of one's actions. Men's sacred objects and ritual paraphenalia are often described as "ngulu", or "frightening and dangerous".'

(3) '"Anger" or "rarru" means to "get wild" or "very angry" because he does not feel sorry for his actions. Pintupi descriptions of "rarru" means that their "ears are closed" and they are not thinking of the consequences of their actions.'

(4) 'When a relative comes to visit the Pintupi say one becomes happy, or rejoices, "pukularrinpa". Or men bringing back meat make people happy. It is the opposite of anger, fighting, and sorrow, and is tied to a major Pintupi image of sociality and cooperation, says Meyers. Ceremonies that include song and dance make people happy, and "pukulpa" was traditionally used to stop fighting with an approaching enemy.'

Another way in which indigenous language use is being preserved in Australia today is in bilingual education programs in schools. Pintupi/Luritja is being taught bilingually in Walungurra School, Northern Territory, since 1983, with Luritja language being taught at Papunya (Black 1983:207–20). Another place of bilingual education is of the Diwurruwurru-jarru Community, where the Walpiri language, (also a Pama-Ngungan language) is being taught at Lajamanu, Yuendumu, Nyirripi, and Willowra schools, near Katherine in the north of the Northern Territory (Bavin 1993:85). Kriol is being taught at Barunga School, also in this area (Rhydwen 1993:155–8). In some of these communities, linguist/teachers work alongside indigenous women while the children are taught traditional ways of food gathering. The names

of 'bush tucker' (food), including plant and animal names, and names for 'body parts, are taught in Kriol and Walpiri Language by the Indigenous women in communities near Katherine, and the linguist/teacher teaches the same names in English.

In other bilingual teaching programs, everyday classes are taught in both traditional indigenous language, and English. A significant problem that arises because indigenous language is an oral not written language, is that new words have to be invented to explain how to write things such as commas and full stops, and also in the teaching of mathematics. This is how and where linguists and teachers are working together with indigenous people to preserve Australia's original languages. It is especially important that this work be done as quickly as possible so that indigenous languages are not only lost to their original users, but to the world as a whole as globalization and the common use of English as the predominant language take over.

References

Amery, Rob. 1986. Languages in contact: the case of Kintore and Papunya (with particular reference to the Health Domain). Language in Aboriginal Australia 1.

Bavin, Edith. 1993. Language and culture: socialisation in a Walpiri community. In Walsh and Yallop.

Black, Paul. 1983. Aboriginal languages of the Northern Territory. Batchelor, NT: School of Australian Linguistics, Darwin Community College.

Blake, Barry. 1991. The handbook of Australian languages. Cambridge: Oxford University Press.

Devlin, Bruce. 1990. Some issues related to vernacular language maintenance: a Northern Territory view. ed. by, C. Walton and W. Eggington. Language maintenance, power and education in Australian Aboriginal contexts, Darwin: Northern Territory University Press.

Dixon, R.M.W. 1980. The languages of Australia. Cambridge: Cambridge University Press.

Eades, Diana. 1993. Language and the law: White Australia v Nancy. In Walsh and Yallop.

Haebich, Anna. 2000. Broken circles: fragmenting indigenous families 1800–2000. Fremantle, Western Australia: Fremantle Arts Centre Press.

Hansen, Kenneth. 1977. Pintupi/Luritja dictionary, 2nd edition. Alice Springs, Northern Territory: Summer Institute of Linguistics, Australian Aboriginals Branch.

Hansen, Kenneth C. 1980. Translating for the Pintupi. Nungalinya Occasional Bulletin No 22. Darwin, Northern Territory: Nungalinya Aboriginal and Torres Strait Islander Church Training and Research Centre, University of the Northern Territory.

House of Representatives Standing Committee on Aboriginal and Torres Strait Islander Affairs. June 1992. Language and culture — a matter of survival. Report of the Inquiry into Aboriginal and Torres Strait Islander Language Maintenance. Australian Government Publishing Service.

Meyers, Fred R. 1976. Cultural concepts of the emotions. From: to have and to hold: A study of persistence and change in Pintupi social life. Ph.D. thesis. Bryn Mawr.

Myers, Fred R., 1986. Pintupi country, Pintupi self: sentiment, place and politics among Western Desert Aborigines. Washington, DC: Smithsonian.

Rhydwen, Mari. 1993. Kriol:the creation of a written language and a tool of colonization. In Walsh and Yallop.

Schmidt, Annette. 1990. The loss of Australia's Aboriginal language heritage. Canberra: Aboriginal Studies Press.

Simpson, Jane. 1993. Making dictionaries. In Walsh and Yallop.

Tindale, Norman. 1974. Aboriginal tribes of Australia. Los Angeles: University of California Press.

Trudgill, Peter. 1995. Sociolinguistics : an introduction to language and society. England: Penguin Books.

Walsh, Michael, and Yallop, Colin, (eds.) 1993. Language and culture in Aboriginal Australia. Canberra, Australia: Aboriginal Studies Press.

Wierzbicka, Anna. 1997. Understanding cultures through their key words. New York: Oxford University Press.

JOANNE MARIE BAUMGARTNER

Pitcairnese

Pitcairnese is the name of one of two closely related varieties of a language initially formed on Pitcairn island and subsequently transplanted to Norfolk island by descendants of the nine English-speaking sailors who mutinied on HMS *Bounty* in 1789 and their Polynesian companions from Tahiti and Tubuai, one of the Austral Islands of the Tahiti group. Linguists use the name 'Pitcairn-Norfolk' (PN) to refer to both, but speakers call the variety spoken on Pitcairn 'Pitkern' and that on Norfolk Island 'Norfolk.' In 1996, Pitcairn Island Council, prompted by the concern that Pitkern might die out because none of the children were speaking it spontaneously, declared it an official language of the island.

Although there is some disagreement about whether PN is a creole rather than a variety of English, the conditions under which it developed were clearly exceptional because we know the place and time of its formation, as well as details of the origins of the primary individuals involved. PN is one of the few cases in which it is possible to directly trace the survival of some British dialectal features to some of the sailors.

The *Bounty*, sent to Tahiti to transport breadfruit trees to the West Indies as a source of food for slaves, contained no pressganged men and was probably the first British ship to sail with an all-volunteer, hand-picked crew of 45. The cliché 'motley crew' well describes the mutineers, for they covered not only a wide social spectrum ranging from their leader, Fletcher Christian, master's mate, born of an aristocratic family originally from the Isle of Man, to common seaman John Adams, a Cockney and orphaned son of a Thames waterman, brought up in a poor house, but they also covered a wide range of regional varieties of English. Four of the nine mutineers were English, two Scottish, one American (from Philadelphia), one part-West Indian from St. Kitts in the Caribbean, a well-educated son of a navy captain (Edward 'Ned' Young), and one, John Williams, assistant armorer, was a native of Guernsey and spoke French, as Captain William Bligh noted in his list of identifying traits of his men. Of Peter Heywood, he remarked that he spoke with the 'Manx, or Isle of Man' accent.

In their search for a place to hide from the British navy, the mutineers initially sailed back to Tahiti and to Tubuai, where they picked up six men and 12 women, one with a daughter, before settling in 1790 on a remote Pitcairn island more than 2,000 kilometers from Tahiti. This small, two- by one-mile, harborless island with its steep cliffs and lack of reefs or anchorage was at the time uninhabited and incorrectly charted on contemporary maps. Pitcairners' first contact with the outside world occurred 18 years after settlement, in 1808, when Captain Folger of the *Topaz* from Boston was greeted in English by Thursday October Christian, Fletcher Christian's son, and was asked whether he knew Captain Bligh.

Apart from this brief encounter, the settlers lived there for the first 33 years in almost complete isolation, thus permitting no outside influences during the formative period of the language. It was not, however, a tranquil time by any means, with nearly 90% of the founding male population murdered by 1800. Disputes over land and the women led to the violent deaths of all the men except four Polynesians and four mutineers, one Englishman (John Adams), a Scot (William McCoy), one Cornishman (Matthew Quintal), and one West Indian from St. Kitts (Edward Young). Young, McCoy, and Quintal died by 1800, leaving only John Adams, then 33 years old, who remained with nine surviving women and the 23 children born on the island. Outsiders did not enter the colony until the 1820s. The first separation in a community that had lived together as a family for more than 60 years occurred in 1856, when the 194 Pitcairners were resettled on Norfolk Island, originally a penal colony. Eighteen months after their arrival, 17 members of the

Young family sailed the 3,700 miles back to Pitcairn. Five years later, a second party returned. Today, fewer than 50 speakers remain on Pitcairn Island. Present-day Norfolk is spoken by about 700 people; the majority of the island's population is not descended from the original Pitcairners. Some 700 other descendants of the original population live in Australia, New Zealand, and elsewhere.

Because the mutineers were outnumbered by two to one, the odds were that Pitcairn would become predominantly Polynesian in language and culture, particularly because all the women were Polynesian (most of them from Tahiti) and would have been the primary caregivers to the children. There are also hints in both boatswain's mate James Morrison's and midshipman Peter Heywood's accounts of the mutiny that the sailors had begun to intersperse Tahitian words in their speech to one another. Midshipman Roger Byam recalls that by the time the *Bounty* sailed with the by now heavily tattooed sailors after a stay of more than five months in Tahiti, all the men knew some Tahitian words. Some were said to be quite fluent and could be heard carrying on conversations with one another in which barely a word of English was spoken. Yet, the language of power on Pitcairn island was clearly English, and many linguists argue that PN was not a fully developed pidgin or creole because there was no point in its history in which English was not the primary language of the community. Stable pidgins rarely develop where contact involves only two languages. Moreover, the descendants seem to identify more strongly with their English rather than Tahitian lineage, and the predominant influence on the vocabulary, pronunciation, and sentence structure of PN is English.

The main influence of Tahitian survives in the form of a large element of Tahitian vocabulary, particularly for flora, fauna, and foodstuffs, e.g. *buhe*, 'sea eel' (Tahitian *puhi*), and *miti*, 'coconut cream sauce' (Tahitian *miti*). Tahitian influence can also be seen in certain semantic (meaning) shifts that have affected English words, such as *hand*, meaning both 'hand' and 'arm' in PN, paralleling Tahitian *rima*. Otherwise, most of the vocabulary is ultimately of English origin, including a number of dialects. Words can sometimes be traced to the dialects of individual mutineers, e.g. the negative imperative ('do not') *dune* from the Scots probably spoken by John Mills, gunner's mate from Aberdeen, or William McCoy of Rosshire, or *moge*, 'thin' (and possibly also the negative form *kaa/kannt*, 'cannot'), from Edward Young's West Indian English. Other Scots forms include *bole*, 'to make a small hole in anything', *devil's needle*, 'dragonfly', *gaggle*, 'to cackle', and possibly *tayte*, 'potato'. There are also archaisms such as *dub*, 'to square and smooth (timber)', *tardy*, 'late', *paunch*, 'stomach', and a few words of American origin, such as *corn*, 'maize',

and *candy*, 'sweets', possibly because of the presence of Isaac Martin, a mutineer from Philadelphia, or because of the influence of American whalers who visited the island. Other dialect forms include *dunnekin*, 'lavatory', and *grub*, 'food'.

A variety of terms with original nautical reference, such as *all hands*, 'everyone', *deck*, 'floor', and *flog*, 'spank', have extended their meanings in PN. Similarly, *heave ho*, a cry of sailors in raising the anchor up, survives in PN *heway me* (*heave me away*) 'to lift up/ throw/take away/ remove', as does *heway* ('heave away'), now obsolete in English. However, this once nautical usage with the meaning 'to fling, throw or cast, to haul up or raise by means of a rope' does exist in dialectal usage in Scots, where *heave away* means 'to throw away', as well as in some English dialects in which *heave* means 'to lift/raise'.

Although PN was formed independently of Pacific Jargon English, an English-based foreigner talk used in encounters between Europeans and Pacific islanders and found on almost all islands in Polynesia and Micronesia by the 1830s because of the whaling industry, it shares some features in common with other Pacific pidgins and creoles descended from it, e.g. bimorphemic (i.e. two-part) question words such as *whatawe*, 'how' (English 'what way'), verbs generally unmarked for tense (past, present) and aspect (whether the action is completed or not), use of *dem* (English 'them') as a plural marker, e.g. *dem ai*, 'the eyes', and absence of the copula 'be', e.g. *whatawe hem*, 'how is he?' The sailors probably knew some rudiments of Pacific Jargon English with adaptations made during their stay at Tahiti. These features could also have been acquired through contact with visiting whalers. Between 1813 and 1852, 400 visits of ships are recorded, three quarters of which were American, mostly whaling ships. For a time, there was a steady market for the islanders' vegetables. Norfolk islanders also had contact with the whaling trade in the late nineteenth and early twentieth centuries. Many worked as crewmen on whaling ships. They were also in contact with the varieties of Melanesian pidgin spoken by Solomon Islanders and New Hebrideans. In 1867, the Melanesian Mission established a training school for Solomon Islanders on Norfolk. From the 1890s until World War II, Norfolk Islanders worked as plantation managers and cadets in the Solomons. Otherwise, however, PN has far more in common with the Caribbean creoles, no doubt attributable to the influence of Edward Young from St. Kitts. The possessive construction makes use of English 'for' as in *dem ai fo yoen*, 'your eyes'.

A few grammatical features, particularly the personal pronouns, reflect Tahitian influence, e.g. the first-person inclusive dual *hemi* ('you and I'), a literal translation of Tahitian *taua*, composed of English 'thou' + 'me', with the initial sound of 'thee' changed to /h/. This pronoun contrasts with *uklun* ('we/us plural'), whose etymology is unclear. PN, like Tahitian, has dual marking in forms such as *yu tu*, 'you two', contrasting with *yoli*, 'you plural', i.e. three or more. This may also reflect archaic English 'all ye' or 'you all'.

The sound system of PN also shows some influence from Tahitian. Lack of a significant distinction between /r/ and /l/, a feature common to many Polynesian languages, including Tahitian, is found in PN forms such as *stolly*, 'lie' (English 'story'), *morla*, 'tomorrow', and *tomolla*, 'day after tomorrow' (both from English '(to)morrow'). This feature (along with variation between /v/ and /w/, also similar to Tahitian) is a feature typical of St. Kitts too, which may suggest additional influence from the speech of Edward Young. Unlike many pidgins and creoles, PN has preserved consonant clusters under some circumstances, whereas Tahitian has none. However, final t/d are often absent in PN, a common feature of colloquial and dialectal English. PN pronunciations such as /flaid/ for 'Floyd' and /kloi/ for 'cry' may reflect a sound shift that probably originated in London and was well under way by the mid-nineteenth century. It is possible that John Adams is the origin of this feature in PN, although it could also have arisen later by independent innovation or contact with other varieties, such as the one spoken by Young. The vowels of words such as *price* and *choice* still rhyme in St. Kitts today in many people's usage. The /ie/ diphthong (double vowel sound) of *gate* could reflect West Indian or English dialectal influence. Despite the presence of two Scots among the founding English speakers, PN has no trace of /r/ after vowels in words such as *cart*, *barn*, etc.

References

Buffett, Alice, and Donald C. Laycock. 1988. Speak Norfolk today. Norfolk Island: Himii.

Dening, Greg. 1992. Mr. Bligh's bad language. Passion, power and theatre on the bounty. Cambridge: Cambridge University Press.

Harrison, Shirley. 1972. The language of Norfolk Island. B.A. Thesis. Sydney: Macquarie University.

Harrison, Shirley. 1985. The social setting of Norfolk speech. English World-Wide 6(1). 131–53.

Laycock, Donald C. 1990. The status of Pitcairn-Norfolk: creole, dialect, or cant. Status and function of languages and varieties, ed. by Ulrich Ammon. Berlin: de Gruyter.

Mackaness, George. 1938. A book of the 'Bounty' and selections from Bligh's writings. London: J.M. Dent & Sons, Ltd.

Ross, A.S.C., and Moverley, A.W. 1964. The Pitcairnese language. New York: Oxford University Press.

Suzanne Romaine

See also **Pidgins and Creoles**

Place of Articulation

Consonants are formed by creating a constriction in the vocal tract; in producing a consonant, the place of articulation is the part of the vocal tract having the greatest constriction. The symbols used for the description of sounds are taken from the IPA chart, which lists the International Phonetic Alphabet.

The major constriction, which may be a complete or partial closure, is made by the articulators coming together, generally by a lower articulator moving toward an upper articulator. The lower articulators are elements of the lower jaw—the lower lip, the lower teeth, and the tongue. The upper articulators are the upper lip, the upper teeth, the palate, the velum, the uvula, and the rear wall of the pharynx. Places of articulation usually have a compound name giving the lower and upper articulators, with the name of the lower articulator first. Thus, apico-dental indicates that

the lower articulator is the apex of the tongue and that the upper articulator is the upper teeth. Occasionally, when the lower articulator is obvious or unimportant, only the upper articulator is named: e.g. *velar*, used alone, is interpreted as meaning dorso-velar. The major places of articulation are shown in Figure 1.

Places of articulation can be divided into four general categories: labial, coronal, dorsal, and guttural. At each place of articulation, sounds with different manner of articulation can be produced: stops have complete closure at the indicated place, whereas fricatives are pronounced with partial closure, which produces a hissing sound.

Labial sounds are made with one or both lips and include bilabials, labio-dentals, and linguo-labials. The lower lip articulates with the upper lip to form a bilabial consonant. The term bilabial is used rather
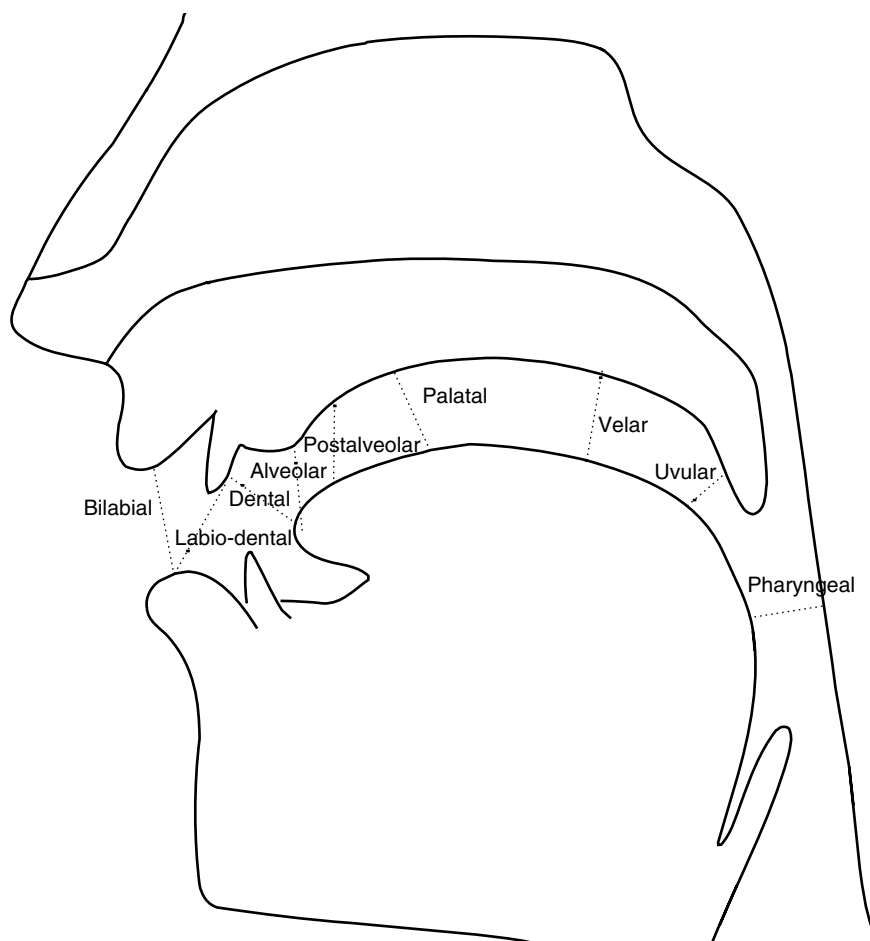
Figure 1

than labio-labial. Bilabial stops [p, b] and the nasal [m] are extremely common in languages; indeed, a language without them (e.g. some Oto-Manuean languages) is noteworthy. The bilabial fricatives [ɸ, β], however, are rather uncommon; languages tend to have bilabial stops and labio-dental fricatives.

The lower lip articulates against the upper teeth to form labio-dental consonants. Labio-dental fricatives [f, v] are very common. Labio-dental stops do not occur as distinctive sounds in any language. The labio-dental nasal [ɱ] occurs as a variation of [m] in English words such as *symphony* ['sɪɱfəni], in which it takes the place of articulation of the following labio-dental fricative.

Linguo-labial sounds are very rare sounds made with the tip or blade of the tongue articulating with the upper lip. The IPA uses a diacritic [͜] underneath a labial symbol for this place of articulation: [p̼ b̼].

Coronal sounds are made with the apex, or lamina, of the tongue. They include dentals, alveolars, alveolo-palatals, postalveolars, and retroflex sounds.

Dental sounds can be made with either the tip of the tongue (apico-dental) or with the blade (lamino-dental). The dental fricatives are voiceless [θ] and voiced [ð]; these sounds both occur in English as in the words *thin* and *then*, respectively. Other dental sounds are shown by using the alveolar symbol with the diacritic [̪]; e.g. [t̪ d̪ n̪ l̪ r̪]. Lamino-dental fricatives [s̪ z̪] can be made with the blade of the tongue near the back of the upper teeth, in addition to the apico-dental fricatives [θ ð]. Danish has a dental approximant [ð̞]. Interdentals are made by thrusting the tongue slightly forward so that the tip protrudes between the teeth; they can be symbolized as [θ̟ ð̟]. In English, apico-dentals are normal, although some people occasionally use interdentals in emphatic or exceptionally careful speech.

Like the dentals, alveolar sounds may be made with either the tip or blade of the tongue, known accordingly as apico-alveolar or lamino-alveolar. The symbols for either are [t d n s z]; all of these sounds occur in English. Alveolar and dental sounds are both extremely common in languages; it is rare, however, for a language to have both, except for fricatives.

Postalveolar sounds are made with the blade of the tongue articulating with the area at the back of the alveolar ridge. The fricatives [ʃ ʒ] and affricates [tʃ dʒ] are quite common in the languages of the world. English has these sounds in *shin* [ʃ], *leisure* [ʒ], *chin* [tʃ], and *gin* [dʒ].

The alveolar fricatives [s z] and the postalveolar fricatives [ʃ ʒ] involve some difficulty in describing their point of articulation accurately. Different speakers hold their tongues in somewhat different positions. With the alveolars [s z], the airstream hits the back of the upper teeth. This causes a turbulence that gives them a distinctive sound. For [ʃ ʒ], the air stream hits the back of the lower teeth and produces a slightly different sound.

Two fricatives are given in the International Phonetic Alphabet as alveolo-palatal; these are different from the postalveolars. They are produced with the tip of the tongue near the lower teeth and with the blade quite close to the back of upper teeth. The symbols are [ɕ] for the voiceless fricative and [ʑ] for the voiced fricative. These sounds occur in Mandarin Chinese and in Polish.

For retroflex consonants, the tip of the tongue curls back, and the underside of the tip articulates with the area at the back of the alveolar ridge. The body of the tongue is quite concave. Retroflex consonants are symbolized as [ʈ ɖ ɳ ʂ ʐ ɭ]; they use the alveolar symbols modified by a lower hook. These sounds are common in languages of India.

The English *r* sound is usually called retroflex; however, in practice, the place of articulation may vary from alveolar [ɹ] to retroflex [ɻ]. Further, many speakers of English use a bunched r, in which the tip of the tongue is near the lower teeth and the body of the tongue is pulled up and back toward the palate. Although the articulations of the retroflex and bunched r are very different, the acoustic effect is quite similar.

Dorsal sounds include those made with the front or back of the tongue: palatal, velar, and uvular.

With palatal sounds, the front of the tongue articulates with the palate. The tip of the tongue points down, often touching the lower teeth. Italian has a palatal lateral [ʎ], as in *figlio*, and a palatal nasal [ɲ], as in *signore*; German has a voiceless palatal fricative [ç], as in *ich*. The palatal stops [c ɟ] and the voiced fricative [ʝ] are not common.

Dorso-velar sounds are made with the dorsum of the tongue articulating with the velum. The dorso-velar stops are [k g] as in English *could, good*; the dorso-velar nasal is [ŋ] as in English *sing*. The fricatives are [x], as in German *Bach,* and [ɣ], the voiced equivalent. Frequently, these sounds are simply called dorsals.

The uvular consonants are made with the dorsum of the tongue articulating with the uvula. They are like velar sounds made very far back in the mouth. The symbols are [q] and [ɢ] for the stops, [ɴ] for the nasal, and [χ] and [ʁ] for the fricatives. Uvular sounds are found in Arabic and Inuktitut. Some French accents have an approximant [ʁ] or fricative [ʁ] for the *r* sound.

The guttural sounds include the pharyngeal, epiglottal, and laryngeal places of articulation.

Pharyngeal consonants are made by moving the root of the tongue backward toward the pharyngeal wall. The root of the tongue is the vertical part, forming the forward wall of the pharyngeal cavity. Many people cannot make a complete pharyngeal stop. A pharyngeal nasal stop is physiologically impossible

because, if no air flows through the pharynx, then no air can flow through the nasal passage. For the fricatives, the symbols are [ħ] and [ʕ]; these sounds are found in Arabic and other Semitic languages.

Epiglottal sounds are extremely rare, but they occur in some Semitic and Caucasian languages. For them, the epiglottis moves backward to articulate with the pharyngeal wall. The symbols are [ʡ] for the stop and [ʜ ʢ] for the voiceless and voiced fricatives.

Laryngeal sounds involve the vocal folds; they include the *glottal stop* [ʔ] and the fricatives [h] and [ɦ]. From a phonological point of view, these sounds can all be thought of as consonants with a laryngeal place of articulation, with the *glottal stop* as a laryngeal stop and the sounds [h] and [ɦ] as laryngeal fricatives. From a purely phonetic point of view, however, each of these could be categorized differently: [ʔ] as a state of the glottis involving complete closure of the vocal folds, [h] as a voiceless vowel, and [ɦ] as a breathy voiced vowel. Glottal stop and [h] are very common; [ɦ] is somewhat rare, although it occurs between vowels in English, as in *ahead*.

In general, the major categories of place of articulation (labial, coronal, dorsal, and guttural) are independent of each other. Thus, it is possible to make two stops at the same time, e.g. a dorsal [k] and a labial [p]. Such a sound is said to have a double articulation; for example, with [kp], the place of articulation is labial-velar. It is symbolized as [k͡p], with the tie-bar showing that the [k] and the [p] are made simultaneously. English [w] is a doubly articulated labial-velar approximant: the lips are rounded and the dorsum of the tongue is partially raised toward the velum. Many English speakers pronounce a double articulation in *kitten* [ˈkɪʔtn̩]. Swedish has a doubly articulated voiceless postalveolar-velar fricative [ɧ]; this is the same as a simultaneous [ʃ] and [x].

Secondary articulations involve the addition of a lesser, or secondary, articulation to a greater, or primary, articulation. Thus, by adding lip-rounding to a [t], a labialized [tʷ] is produced. The primary place of articulation is alveolar; the secondary articulation is labial. The common secondary articulations are (shown with [t]) labialization [tʷ], palatalization [tʲ], velarization [tˠ], and pharyngealization [tˤ]. The IPA has alternative symbols for either velarization or pharyngealization [ɫ].

Labialized sounds involve lip rounding; they are quite common. Palatalized consonants have the front of the tongue more toward the palatal region than usual. Palatalized consonants are common in some Slavic languages. Velarized consonants have the tongue toward a position that is more velar than usual, i.e. a high back unrounded vocalic position. The English dark [ɫ], occurring at the ends of words such as *ball*, is an example of a velarized lateral; here, the dorsum of the tongue is raised slightly toward the velum. Pharyngealized sounds are found in Arabic; they involve a retraction of the root, thus effecting a narrowing of the pharynx.

References

Catford, J.C. 1977. Fundamental problems in phonetics. Bloomington, IN: Indiana University Press.

Catford, J.C. 1988. A practical introduction to phonetics. Oxford: Oxford University Press.

Gimson, A.C. 1980. An introduction to the pronunciation of English. London: Edward Arnold.

Hardcastle, William J., and John Laver (eds.) 1997. The handbook of phonetic sciences. Oxford: Blackwell.

International Phonetic Association. 1999. Handbook of the International Phonetic Association. Cambridge: Cambridge University Press.

Ladefoged, Peter. 1993. A course in phonetics. Ft. Worth, TX: Harcourt Brace and Jovanovich.

Ladefoged, Peter. 2000. Vowels and consonants: an introduction to the sounds of languages. Oxford: Blackwell.

Ladefoged, Peter, and Ian Maddieson. 1996. Sounds of the world's languages. Oxford: Blackwell.

Laver, John. 1994. Principles of phonetics. Cambridge: Cambridge University Press.

Minifie, Fred, Thomas J. Hixon, and Frederick Hixon. 1973. Normal aspects of speech, hearing, and language. Englewood Cliffs, NJ: Prentice-Hall.

Pullum, G.K., and W.A. Ladusaw. 1996. Phonetic symbol guide. Chicago: University of Chicago Press.

Rogers, Henry. 2000. The sounds of language: an introduction to phonetics. Harlow, England: Pearson.

HENRY ROGERS

See also **Anatomy of the Articulatory System**

Plurality

The majority of languages have singular and plural forms of words to indicate a single object or a set of more than one object. Some languages (Philippine languages, some languages of Southern Asia, Baltic languages) have a dual form of words to indicate a set of two objects. Dual nouns are used in archaic dialects of

Lithuanian, and some relics are still known in Russian, Ukrainian, and some other Slavic languages. For example, in Russian the word *eye* has the form: *глаз* [glaz] (singular), *глаза* [glaza] (plural). The plural form is actually a relic of the dual. The origin of the dual in these languages can be explained by the fact that some things exist only in a pair: like ears, eyes, knees, legs, and sides.

In most of the modern languages, the dual form has disappeared. Only two categories, i.e. the singular and plural (more than one object) are used; for example, the singular word *leg* in English takes the plural form of *legs*; in Russian it would be *нога* [noga]—*ноги* [nogi]; and in Turkish the same word would be *bacak* in singular and *bacaklar* in plural.

There are also languages that do not have a specific form to indicate plural objects. This can be illustrated by examples from Malay, where a noun *mata* (eye) would normally imply a plural meaning (there are usually two eyes on a face). The same principle would apply to *kaki* (legs). When necessary, the plural form in Malay is created by repeating the noun—*kamus-kamus* (dictionaries), *kanak-kanak* (children).

The plurality of nouns and pronouns as objects in the sentence affects subject–verb agreement. For example, the singular English pronoun *I* requires a singular verb—*I have a beautiful vase for these flowers.* While the plural form of *I*—*we* requires a plural verb—*We have a beautiful vase for these flowers.* The singular noun *sister* requires a singular verb—*My sister goes to the movies every Sunday.* The plural *sisters* requires a plural verb—*My sisters go to the movies every Sunday.*

The following English nouns are grouped according to objects they refer to:

(1) *Sister, house, table, book, mother, minister, teacher, doctor;*
(2) *Milk, coal, chocolate, mud, water, gold, information, sugar, cream;*
(3) *Love, hatred, friendship, hope, charity, justice;*
(4) *Family, group, committee, class, parliament.*

The first group refers to concrete, tangible objects, or persons. They are *count nouns.* The second and third groups of nouns indicate substance (2) or abstract (3) nouns that are not discrete, and therefore cannot be counted. They are *mass nouns.* The last group represents entities or bodies that consist of a number of members. These nouns can be counted. They are *collective nouns.*

Count Nouns

Count nouns can be preceded by the word *some* or the articles *a, an*, and *the.* For example: *A tree has fallen during the storm. The tree has fallen during the storm. Some trees have fallen during the storm.* They can have both singular and plural forms. Most of the count nouns form their plurals by adding -(*e*)*s*. However, there are a number of irregular words, and words borrowed from other languages, from Greek and Latin in particular, that create their plurals in specific ways (see examples below). There are also some English words that can only be either singular or plural.

The plural form of count nouns is formed by adding -(*e*)*s* to the noun:

flower	*flowers*
lady	*ladies*
balloon	*balloons*
lunch	*lunches*
bush	*bushes*
fairy	*fairies*
potato	*potatoes*

A number of irregular nouns in English form their plurals differently, i.e. by changing their vowel or by adding *-en*.

man	*men*
woman	*women*
foot	*feet*
goose	*geese*
mouse	*mice*
louse	*lice*
child	*children*
ox	*oxen*

Those count nouns, which end with *-f*, form plurals by changing *-f* to *-ves*, for example,

wolf	*wolves*
wife	*wives*
scarf	*scarves*

Words borrowed from other languages, Latin and Greek in particular, create their plurals according to the foreign (original) rule, for example:

datum	*data*
stratum	*strata*
larva	*larvae*
criterion	*criteria*

Some English nouns are used only in their singular form, such as *music, chemistry, physics, ceramics, mathematics*, etc. At the same time, there are nouns that are only used in the plural: *jeans, trousers, scissors, spectacles, measles*, etc. However, the language allows for the following phrases if there is a need to indicate the number of some of the above-mentioned words. For example—*Tanya knew she desperately needed another pair of jeans. The office assistant has ordered two pairs of scissors for the department.*

There are cases of nouns that do not change for plural, e.g. *sheep, deer, fish, salmon*, etc. Also, there are words that have two different plural forms that convey different meanings, such as:

craft:	*craft (boats)*
	crafts (handicrafts)
index:	*indexes (alphabetical lists of names, subjects, etc.)*
	indices (algebraic exponents)
medium:	*mediums (at spiritual séances)*
	media (newspapers, television)

Collective Nouns

Collective nouns are the nouns that indicate a group, a body, or an organization that consists of a number of members, for example *family, class, parliament, committee*. They form their plurals in the same way the count nouns do by adding (*e*)*s: families, classes, parliaments, committees*. The fact that a collective noun refers to a number of people does not affect the subject–verb agreement in a sentence. Such words in their singular form are used with singular verb, e.g.—*My family enjoys skiing. The Social Club Committee usually plans the recreational activities of the employees.*

Mass Nouns

Mass nouns are not usually used with the definite article; for instance—*Loyalty is his most valuable characteristic. Sugar adds sweetness to any drink.*

However, in the following examples the mass nouns are modified by further information and can be used with the definite articles—*The love John and Matt shared for their mother finally ended the war between the brothers. The chocolate made in this town is delicious.*

Since the mass nouns refer to abstract, not discrete objects, they cannot be counted. Therefore, they usually do not allow for plural morphology. However, they allow for other means of measurement:

A bottle of milk;
Three cups of tea;
Bars of chocolate;
Patches of mud;
A useful piece of information.

The appearance of such words as *sugars* or *lights* is possible only if they refer to concrete, discrete entities; for example: *Switch off the lights please. Can I have two sugars in my coffee please?*

References

Dummet, Michael. 1993. Grammar and style for examination candidates and others. London: Duckworth.

Haegeman, Liliane, and Jacqueline Gueron. 1999. English grammar and generative perspective. Oxford, UK: Blackwell Publishers.

Horan, Ronald S. 1996. Exploring English: a complete analysis of modern grammar. Leichhaardt, Australia: Honeysett Press.

Hudson, Richard. 1984. Word grammar. Oxford, UK: Blackwell.

Johansson, Stig. 1980. Plural attributive nouns in present-day English. Lund: CWK Gleerup.

Bach, Emmon, et al. (ed.) 1995. Quantification in natural language, Vol. II. Dordrecht: Kluwer Academic Publishing.

Reid, Wallis. 1991. Verb and noun number in English: a functional explanation. London: Longman.

ALFIA ABAZOVA

Polish and West Slavic Languages

The present-day West Slavic language area is made up primarily of the national territories of Poland, the Czech Republic and Slovakia, as well as a small part of south-eastern Germany. By the end of the sixth century CE, migrations of Slavic-speaking people had settled in a wide area of northwestern Europe, and these communities extended as far west as the River Elbe. By the ninth century CE, distinctly West Slavic elements had become noticeable within the Common Slavic used by these settlers, and, by the end of the tenth century, features distinctive of the three West Slavic subdivisions of Lechithic, Czecho-Slovak, and Sorbian had emerged. Apart from one or two isolated remaining pockets of Slavic, the more westerly Slavic-speaking areas were gradually resettled by German speakers in the middle ages, resulting in approximately the West Slavic language area as it exists today.

The modern West Slavic languages are Polish, Czech, Slovak, and Upper and Lower Sorbian. To these is also often added a further member, Cassubian, whose status is disputed between its being a separate language or a dialect of Polish. Until the mid-eighteenth century, another West Slavic language, Polabian, also existed. Polish, Cassubian, and Polabian make up the Lechithic subgroup of West Slavic, Czech, and Slovak make up the Czecho-Slovak subgroup, and Upper and Lower Sorbian make up the Sorbian group. Other divisions have been suggested in the past, involving the merging of Sorbian with either Lechithic or Czecho-Slovak, but this three-group model is the most widely accepted.

West Slavic is distinguished from South and East Slavic primarily on the basis of phonetic differences, such as the retention of /tl/ and /dl/ clusters, which have been reduced to /l/ in South and East Slavic languages. West Slavic languages also have fixed word stress, while most other Slavic languages have moveable stress.

All the West Slavic languages use a Roman-based alphabet, with each language having a set of accented and/or special characters. Polish has the largest set of special characters, including 'hooked' vowel characters that represent nasalized vowels.

Two features of the West Slavic languages have been of particular general interest to linguists. Both are shared with other Slavic languages, but one is especially closely associated with work on West Slavic.

The first feature is the category of aspect, which stands alongside tense to indicate whether an action or event is completed or ongoing, as opposed to whether its occurrence is in the past, present, or future. This is evident in the existence of pairs of related verbs, such as the Lower Sorbian *cytaś* and *docytaś*, where *cytaś* focuses on the general act of reading, while *docytaś* is used for a completed reading session (meaning roughly 'read through to the end').

The other feature is word order within the sentence. All the West Slavic languages mark grammatical functions by the use of special word endings, rather than by word order as is the case in English. They thus permit considerable flexibility in word order, and they use this to give prominence to particular pieces of information within the sentence. This phenomenon has sometimes gone by the name of 'functional sentence perspective' and its theoretical status within general linguistics originated with work on Czech by the Prague School in the 1930s.

Polish and Cassubian

Polish has the largest number of speakers of the West Slavic languages. It is the official language of the Republic of Poland and is thus spoken by around 38.5 million people in that country. It has been estimated that there are several million more speakers outside Poland, including immigrant communities in the United States of America and elsewhere.

The earliest evidence that we have for Polish dates from the twelfth century and consists solely of proper names within an otherwise Latin document. The earliest surviving complete sentence in Polish dates from the thirteenth century and the earliest complete text dates from the fourteenth century.

Within present-day Polish, five main dialect areas are usually recognized: Wielkopolska is an area toward the northwest of Poland, centered around Poznan, Mazowsze is in the northeast and includes Warsaw, Małopolska is in the southeast, Śląsk is a small area toward the southwest, and Kaszuby is another small area in the north around the port of Gdansk. In addition to these, there are also mixed dialect areas, both in the extreme northeast and in the west along the border with Germany. These latter areas were part of Germany until the end of World War II and were then settled by Polish speakers of various dialects.

An increasingly standard form of literary Polish, drawing mostly on the dialects associated with the main cultural centers of Poznan and Gniezno (in the Wielkopolska area) and Cracow (in the Małopolska area) can be identified from the sixteenth century, when there was a flourishing 'golden age' of Polish literature. This literary language was also influenced to a certain extent by Czech, as well as by borrowings from Latin. Following a period of decline for written Polish in the early eighteenth century, a revivalist movement emerged in the late eighteenth and early nineteenth centuries, which aimed to codify the standard language. Although members of this movement took note of the earlier literary language of the sixteenth and seventeenth centuries, it was not primarily an archaizing movement (in contrast to the codification of Czech, outlined below) and consequently the differences between the codified written and contemporary spoken languages were not all that great. Throughout this period, and also thereafter, there was a strong opposition to the adoption of distinct regionalisms into standard Polish. A version of Polish based on the standard written language has become adopted as the standard spoken variety. Today, most Polish speakers who speak a dialect are also able to speak this standard language, and they use it in more formal and public settings.

Distinctive features of Polish include the nasalized vowels mentioned earlier, the marking of gender as well as number and person in certain forms of the verb, and the use of third-person ('he' or 'she' or 'it') verb forms instead of second-person ('you') forms in respectful address.

The Kaszuby dialect (Cassubian in English) is, as stated earlier, often recognized as a separate language within the West Slavic group, although within Poland there is a greater tendency to see it as a dialect of Polish. It certainly possesses greater differences from standard Polish than do the other dialects, and Poles from outside the Cassubian region have difficulty in understanding it. Its status as a separate language is strengthened by a strong Cassubian ethnic identity. It also has its own literary tradition and is used regularly in the local media. Despite these outward attributes of a relatively cohesive language, Cassubian has a high degree of internal dialect variation. There have been various attempts to codify the language, but, to date, there is no formally recognized standard. Arguments against the status of Cassubian as a language are its close relation to Polish and the fact that the Cassubian area is politically a part of the Republic of Poland.

References will sometimes be found to a further West Slavic language, Slovincian, which is now extinct. This was located geographically adjacent to Cassubian, to the northwest of the main Cassubian-speaking area. Although the residents of that area referred to their language as Slovincian, it is almost universally recognized by linguists as having been a subdialect of Cassubian. A combination of earlier Germanization and postwar population movements led to the extinction of Slovincian by the 1950s.

Czech

Czech is the official language of the Czech Republic, spoken by between 9 and 10 million inhabitants, as well as by around 60,000 inhabitants of neighboring Slovakia and by immigrant communities elsewhere.

The earliest surviving example of Czech dates from 1057 CE and consists of a set of glosses on a Latin manuscript. From the thirteenth century onward, a prolific Czech literary tradition exists, which flowered especially in the sixteenth century.

Czech is unusual in that there is a larger than average difference between the standard written and spoken languages. The standard spoken language—known as Common Czech—represents the natural evolution of the Czech language over time. In contrast, standard written Czech stems from essentially artificial codification by linguists in the early to mid-nineteenth century, the most influential of whom were Josef Dobrovský and Josef Jungmann. Dobrovský, who was mainly responsible for codifying the grammar of Czech, based his grammar on substantially earlier literary usage from the golden age of the sixteenth and seventeenth centuries. Thus, in comparison with Common Czech and other West Slavic languages, standard written Czech shows a number of archaic features. For instance, the

six sets of noun endings from earlier Slavic are better preserved in Czech than in any of the other West Slavic languages. Jungmann was more concerned with widening the use of Czech, although his attitude was also essentially purist and he was responsible for introducing a number of distinctively Czech new words in preference to borrowing from other languages.

Besides standard spoken Common Czech, Czech divides into three main dialect areas: Bohemian, Hanák (or Central Moravian), and Lach (or Silesian). Additionally, there are two transitional dialect areas: a southeast Moravian area, whose dialects are closely related to Slovak, and a mixed Czech-Polish dialect area. Standard Common Czech is linked most closely to the central Bohemian dialects, and particularly to the dialect of the main urban center, Prague. There is, however, a degree of regional variation even within Common Czech, and some now suggest that these are of sufficient magnitude to argue that two distinct regional versions of Common Czech exist—Bohemian Common Czech and Moravian Common Czech. As with Polish, there are also mixed dialect regions in areas along the border with Germany, which had been German-speaking prior to the end of World War II.

Slovak

Slovak is the official language of Slovakia, spoken by around 4.6 million people there. A further 300,000 or so speakers are resident in the neighboring Czech Republic.

Slovak is closely related to Czech and the two languages are, broadly speaking, mutually intelligible. This is particularly the case with the southeast Moravian dialects of Czech. However, there are also some differences between the two languages in both sound and word structure. This is especially true of the written languages, owing to the different language varieties that were drawn on during the period of their codification.

As with Czech, the earliest surviving evidence for Slovak comes from a set of eleventh-century glosses. However, in contrast, no substantial literary tradition existed for Slovak prior to around 1780: instead, Czech took on the role of the written language, although Slovak writers did introduce some distinctively Slovak features into their written Czech. A further similarity with Czech is in the date that, and the process by which, the eventual standard written language did develop. Written Slovak was not codified until the 1840s, when a standardized form was produced by L'udevít Štúr in association with a group of other Slovak intellectuals. However, unlike Czech, where much earlier written usage exerted a strong influence on the newly codified language, the Slovak written language was based primarily on a synthesis of the contemporary spoken Central Slovak dialects. This

must at least in part have been a necessity, owing to the paucity of a written Slovak tradition, but it also mirrors the codification of Polish, where contemporary usage had more influence than it did with Czech.

Slovak has three main dialect areas—West, Central, and East—with the West Slovak dialects being closest to their Czech neighbours.

Sorbian (Upper and Lower)

The Sorbs are an indigenous Slavic minority who live in the region of eastern Germany known as Lusatia. Until 1990, Lusatia was part of the former German Democratic Republic. It now belongs to the federal states of Saxony and Brandenburg. In earlier times, Sorbian dialects extended eastward beyond the River Neisse into what is present-day Poland, as well as westward and northward into further regions of Germany. However, Sorbian is now wholly contained within the Lusatian region.

Today, there have been estimated to be around 60,000–70,000 speakers of Sorbian, although all Sorbian speakers are now bilingual in Sorbian and German, and a number of these speakers use the German language for most everyday purposes. Sorbian has survived most strongly in small, closely knit rural communities, and has been noted as being especially strong in a group of villages to the north of Bautzen in Saxony. However, although its strongest base is among the rural community, Sorbian is not a low-prestige rural dialect and has a lively literary and scientific culture.

Apart from proper names, the earliest surviving evidence for Sorbian vocabulary consists of glosses in Latin manuscripts dating from the twelfth century. The earliest continuous Sorbian texts date from the sixteenth century. These earlier Sorbian texts were written in local dialects with distinct differences from one another. In the eighteenth century, three standard written varieties emerged: Lower Sorbian, Catholic Upper Sorbian, and Protestant Upper Sorbian, with Upper Sorbian thus being split on denominational lines. (Since the Lower Sorbians were Protestants, the issue did not arise for that variety.) At that time, only religious texts were published in Sorbian. From the mid-nineteenth century onward, alongside an increase in nonreligious writing, the denominational split in Upper Sorbian was gradually diluted, although it was not until after World War II that the present twofold division of Sorbian was firmly established.

There are two present-day standard varieties of Sorbian. Upper Sorbian is the language of the Sorbs in Upper Lusatia (centered around Bautzen in Saxony) and is the larger of the two varieties with roughly 40,000–50,000 speakers. Lower Sorbian is the language of the Sorbs in Lower Lusatia (centered around Cottbus in Brandenburg), spoken by approximately 20,000 speakers. Upper and Lower Sorbian have a number of differences from one another, and these occur at all language levels, including sound structure (e.g. Lower Sorbian has *góra* in place of Upper Sorbian *hora*), vocabulary (e.g. Lower Sorbian has *jo* for 'yes', whereas Upper Sorbian has *haj*), and grammar (e.g. Upper Sorbian makes more use of possessive adjectives than Lower Sorbian).

In contrast to the two standardized varieties, the features that distinguish the several Sorbian dialects show a rather complex distribution, given the small size of the language area. Nevertheless, broadly speaking, it is possible to make a primary distinction between Upper and Lower Sorbian dialect areas, with a transitional area, showing features of both, located in between.

Having coexisted for such a long time alongside German, Sorbian has been affected by a number of Germanic influences, several of which are not shared by the other West Slavic languages. These influences have occurred not only at the vocabulary level but have also affected the grammar of Sorbian. For example, most Slavic languages do not have articles (such as *a*, *an*, and *the* in English). However, under German influence, Sorbian has made use of its demonstrative pronouns and the number 'one' as, respectively, definite (*the*) and indefinite (*a/an*) articles, although this practice is discouraged in the standard written languages, which have been subject to Slavic purist movements. Lower Sorbian probably retains more German influences than Upper Sorbian, which was subject to a more intensive push toward purism. German influence on Lower Sorbian was formerly also evident at the orthographic level, with much Lower Sorbian publishing using the German *Fraktur* type until around the time of World War II.

As a small minority language within a German-speaking context, the number of speakers of Sorbian has declined substantially over the past century. However, there are lively movements to preserve its use and the future is not necessarily bleak. The rights of the Sorbs have been legally guaranteed within a bilingual framework since the end of World War II, firstly by the former German Democratic Republic and subsequently by the federal states of Brandenburg and Saxony. To promote and maintain the language, a number of initiatives exist, including the provision of bilingual schooling from an early age (the WITAJ initiative).

Polabian

Polabian, sometimes known as Dravaeno-Polabian, is an extinct West Slavic language. The last native speaker is reported to have died in 1756, although there are records of someone able to recite the Lord's Prayer in Polabian as late as 1798. Polabian was formerly spoken in a small

part of the Lüneburg area, which is situated in north Germany between Hamburg and Hannover. It was thus surrounded by German-speaking territory from an early point in its history and was geographically isolated from the other Slavic languages. (Contrast Sorbian, which has retained a border with both Polish and Czech.) Not surprisingly, therefore, it shows evidence of a very strong German influence, especially from the regional Low German varieties, both in vocabulary and grammar.

Very little Polabian material has survived, and much of what has is in the form of isolated words and phrases compiled into dictionaries in the last century of the language's existence. Only a small number of complete texts, such as the Lord's Prayer and a folk song, have been preserved. A figure of around 6,000 recorded vocabulary items has been suggested. This has nevertheless provided enough evidence to suggest that at least three dialects existed.

Despite this lack of material, and in view of its unusual isolation from related languages, Polabian has received continued and substantial attention from linguists, including notable figures such as Trubetzkoy.

References

Auty, Robert. 1960–1961. Dialect, Κοινὴ and tradition in the formation of literary Slovak. Slavonic and East European Review 39.

Auty, Robert. 1976. Problems in the formation and development of the Czech literary language. Slavic linguistics and language teaching, ed. by Thomas F. Magner. Cambridge, MA: Slavica.

Comrie, Bernard, and Greville G Corbett (eds.) 1993. The Slavonic languages. New York: Routledge.

De Bray, R.G.A. 1980. Guide to the West Slavonic languages. Columbus, OH: Slavica.

Olesch, Reinhold. 1989. Gesammelte Aufsätze. 1. Dravaeno polabica. Cologne and Vienna: Böhlau.

Polański, Kazimierz. 1993. Polabian. In Comrie and Corbett.

Rothstein, Robert A. 1993. Polish. In Comrie and Corbett.

Short, David. 1993. Czech. In Comrie and Corbett.

Short, David. 1993. Slovak. In Comrie and Corbett.

Stone, Gerald. 1972. The smallest Slavonic Nation: the Sorbs of Lusatia. London: The Athlone Press.

Stone, Gerald. 1993 Cassubian. In Comrie and Corbett.

Stone, Gerald. 1993. Sorbian. In Comrie and Corbett.

Townsend, Charles E. 1990. A description of spoken Prague Czech. Columbus, OH: Slavica.

ANDREW WILSON

See also **Indo-European 5: Slavic**

Politeness

Politeness refers to how a speaker chooses to phrase an utterance in a particular context, a choice that can reflect the speaker's view of the interpersonal context. There are a number of different approaches to politeness. In a social normative view, to be polite is to speak formally and to behave in accordance with situation-specific rules regarding appropriate linguistic behavior. A variant of this approach has been to regard politeness as one component of pragmatic competence (Leech 1983). In this view, there are conversational maxims (e.g. tact and generosity) that stipulate which linguistic form is preferred over another in any particular context.

Currently, the most popular approach to politeness, based on the work of Brown and Levinson (1987), views politeness as the linguistic means by which people manage each other's 'face'. The concept of face (Goffman 1967) refers to the identity that people present during the course of their interactions with others. Face is assumed to be universal and of two types: negative face, or the desire to have autonomy of action, and positive face, or the desire for closeness with others. One's face is fragile and subject to continued threat during social interaction. For example, criticisms threaten the recipient's positive face, and requests threaten the recipient's negative face. Social interaction thus presents a dilemma for interactants; they want to maintain each other's face, but they are also motivated to perform actions that are face-threatening. This conflict is solved by engaging in face-work, or, more precisely, by being polite.

According to Brown and Levinson (1987), politeness is conveyed by deviating from maximally efficient communication; politeness is thus roughly equated with indirectness. There are many ways this can be accomplished, and Brown and Levinson organized politeness into five superstrategies. They provided evidence for these strategies from three different languages and argued that the linguistic strategies for conveying politeness are universal. The least polite strategy is to perform an act without any politeness, or to perform the act directly. The most polite strategy is to perform the act with an off-record form. Off-record forms are deliberately ambiguous and hence perform a face-threatening act indirectly (e.g. *It's warm in here* as a request to have the window opened). Falling between these two extremes are on-record (i.e. relatively unambiguous) acts with redress emphasizing either positive face or negative face. The former, termed 'positive politeness', functions via an exaggerated emphasis on closeness or solidarity with the hearer. Examples

include the use of slang and familiar address terms, jokes, and presumptuous optimism (*You'll loan me your car, won't you?*). The latter, termed 'negative politeness', functions via attention to the recipient's autonomy. This is accomplished by symbolically lessening any imposition on the hearer. A common method is to use conventionally indirect forms (e.g. *Could you shut the door?*). Research has provided partial support for this typology. It appears, however, that politeness strategies are very diverse and may not be completely captured with this scheme. Also, it is unclear whether politeness should be equated with indirectness; very indirect strategies may sometimes backfire and may be perceived as manipulative.

One of the most important features of a face-management view of politeness is the attempt to specify the relationship between politeness and the interpersonal context. Brown and Levinson (1987) proposed that politeness will vary as a function of the weightiness (or degree of face-threat) of the act to be performed. Weightiness is contextually determined and assumed to be an additive weighting of the following three variables: the intrinsic (and culturally bound) degree of imposition of the act itself (e.g. asking for a loan of $5 is less imposing than asking for a loan of $500), the power of the hearer relative to the speaker, and the degree of social distance between the interlocutors. Thus, increased weightiness (hence, in general, greater politeness) occurs as a function of increasing imposition, hearer power, and relationship distance. Note that in this way the politeness of an utterance cannot be determined in isolation, but rather only in a specific interpersonal context. For example, an utterance perceived as polite when used with an underling may not be polite when directed toward a superior.

Research has provided partial support for these ideas. Many studies have demonstrated that greater speaker power is associated with decreased politeness. Fairly consistent support has also been found for the imposition variable, with increasing imposition associated with increasing politeness. The effects of relationship distance have been the most problematic; whether greater politeness is associated with increasing relationship distance is not clear.

One of the most important issues surrounding politeness is its status as a cultural universal. Although few would argue the claim that politeness exists in all cultures, the claim that positive and negative faces are universal desires motivating the form that politeness takes has been questioned. For example, it has been argued that negative face is relevant only in Western cultures, or cultures in which there is an emphasis on individual autonomy. Although Brown and Levinson (1987) marshaled impressive evidence for the cross-cultural validity of their politeness strategies, empirical support for their claim that they are universal has been rare. Clearly, there is great cultural variability in politeness. The crucial question is whether this variability is a result of differing cultural conceptions of face or whether these differences can be explained at a lower level of abstraction. Face wants may be universal, but cultures will vary in terms of what threatens face, who has power over whom, how much distance is typically assumed, and so on.

Recently, there have been some new developments in politeness theory. Some researchers have used it as a framework for explaining how people interpret indirect utterances. Others have begun to examine nonverbal politeness. Finally, many researchers are now using politeness theory in applied contexts as a means of examining and explaining linguistic behavior in specific contexts, such as courtroom trials, classroom interaction, and political campaigns.

References

Brown, P., and Levinson, S. 1987. Politeness: some universals in language usage. Cambridge, UK: Cambridge University Press.

Fraser, B. 1990. Perspectives on politeness. Journal of Pragmatics 14. 219–36.

Goffman, E. 1967. Interaction ritual: essays on face to face behavior. Garden City, NY: Anchor.

Holtgraves, T. M. 2001. Politeness. Handbook of language and social psychology, 2nd edition, ed. by H. Giles and W.P Robinson. London: Wiley.

Lakoff, R. 1973. The logic of politeness: or, minding your p's and q's. Papers from the ninth regional meeting of the Chicago Linguistic Society, 292–305.

Leech, G. 1983. Principle of pragmatics. London: Longman.

Watts, R. J. 2003. Politeness. Cambridge, UK: Cambridge University Press.

THOMAS M. HOLTGRAVES

Possessives

Possessive marking is used to indicate that two nouns or noun phrases stand in a modifying relationship, namely that of possession. One noun (phrase) is the modifier, the possessor, and the other noun (phrase) is the modified, the possessed. According to the position of the modifier, two kinds of possessives are available

in English: one is marked by the genitive case (*my father's books*) and the other by a preposition (*the books of my father*).

According to traditional grammars, the genitive possessive refers to a specific entity or individual. It is more frequent with modifiers referring to animate entities, i.e. living things, but certain inanimate modifiers can also take the genitive, such as geographical names (**Canada's** *farms*), temporal nouns (**yesterday's** *paper*), and the names of institutions (**the university's** *council*). The use of inanimate modifiers also depends on the modified noun and is generally possible with modified nouns referring to places (*the earth's **surface**, the water's **edge**, at the road's **end***) and measures (*arm's **length***). The preposition *of* is used in partitive constructions (*a piece of wood, a kind of letter*) and expressions denoting a place of origin (*Jesus of Nazareth*), quality (*a man of affairs*), and material (*a sword of steel*).

The genitive modifier can function as a determiner and thus excludes the use of other determiners (such as *the* or *a*). For example, it is not possible to say **the** *my father's **book***. To use a determiner, the so-called 'double genitive' construction must be used—that is, *a book of my father's*. The double genitive can be used with the indefinite article *a* and with demonstrative pronouns such as *that* or *this* (**a/that/this** *book of my father's*). The definite article can generally not be used in double genitive constructions: **the** *book of my father's* is not a possible English phrase. However, if there is a modifying relative clause, use of the definite article becomes possible: **the** *book of my father's **that is on the table***.

Genitive and prepositional markings can reflect the subject/object distinction in constructions like *the enemy's destruction of the city*. Here, the nouns *enemy* and *city* receive the same interpretation as in the sentence *the enemy destroys the city*. In both cases, *the enemy* has a subject-like function, while *the city* is the object. This parallelism was an important piece of evidence in a heated linguistic debate concerning nominalization, the process of deriving a noun such as *destruction* from the corresponding verb, here *destroy*. In grammatical theories of the 1960s, it was assumed that a noun phrase like *John's presentation of the project* is derived from the synonymous sentence *John has presented the project*.

This view changed with Noam Chomsky's widely quoted article 'Remarks on nominalization' (1970). Chomsky argued—convincingly to many linguists—that any noun can 'project' structures containing a subject and an object. Thus, for phrases like *the enemy's destruction of the city*, the base structure would be [*enemy destruction city*]. For comparison, the base structure for *The enemy destroyed the city* would be [*enemy destroy city*]. Nouns and verbs thus 'generate' or 'project' very similar structures at an abstract level.

The only difference between noun phrases and verb phrases (or sentences, for that matter) lies in the elements that are used to mark the exact relationships between the individual items. In the case of noun phrases, the possessive is the relevant grammatical marker for the subject. The markers *'s* and *of* thus were considered to be 'semantically null', i.e. they do not contribute any particular meaning to the sentence. Chomsky (1986) assumed that both are simply overt realizations of the genitive (case).

In fact, possessive markers have no semantic content in cases of so-called 'inalienable' possession. For example, the expression *John's picture* may have two 'inalienable' readings: (1) 'John has created the picture' vs. (2) 'John is in the picture'. Both are cases of inalienable possession, because nobody can take away the fact that 'John created the picture', or nobody can take John out of the picture (without destroying the picture). In other words, the entities *John* and *picture* are inseparable and presuppose each other (i.e. an inherent relationship). If *John* is construed as a creator of the picture, this noun is inserted in the subject position [*John picture*], which automatically leads to the genitive expression *John's picture*. If, on the other hand, *John* is the object of *picture*, the base structure would be [*picture John*], and the default phrasing would be *the picture of John*. However, if no creator is specified, i.e. if the subject position is left empty in the base structure, English allows 'promotion' of the object, which again yields the genitive *John's picture*. Hence the potential ambiguity of English genitive expressions.

However, Mona Anderson (1983) argued that the genitive *'s* does contribute independent meaning in cases of 'alienable' or 'true' possession. Thus, the expression *John's picture* can also be interpreted as 'The picture that John owns'. This means that the relation between *John* and *picture* is not inherent but is established by means of the 'word' or 'lexical unit' *'s*. Note that *'s* can be labeled a word, since it fulfills a function similar to 'full' words, namely *possess* or *own*. In other words, in this case the phrase *John's picture* is projected not from the noun *picture* but rather from the possessive *'s*, and the base structure would be [*John 's picture*], similar to [*John possess picture*] or [*John own picture*].

The idea of lexical *'s* has been commonly accepted since the second half of the 1980s, and it has been extended to other structures. For instance, *his book* has been analyzed as deriving from the base structure [*he's book*], where *he* is the subject and *book* is the object of the word *'s*.

In a crosslinguistic perspective, the distinction between genitive and prepositional possessives is not universal. In French, for example, a genitive marker analogous to the English *'s* does not exist. Thus, the

English expression *Robert's friends* is translated in French as *les amis de Robert* (literally *the friends of Robert*). Russian, on the other hand, has no preposition corresponding to the English *of*. For instance, two constructions are available to translate the English expression *Robert's friends*: (1) *Robertovy druzja*, where *-ov-* is a possessive marker, and (2) *druzja Roberta*, where *-a* is a morphological manifestation of genitive case.

Historically, the marker *'s* originates from the genitive case singular marker *-es* used in Old English. For example, in the sentence *Him sceamode þæs mannes* 'He felt ashamed for the man', this ending is observed on the noun stem *mann* 'man' preceded by the genitive form of the demonstrative *se* 'the, that', i.e. the literal translation would be something like 'To-him there-was-shame the-of man-of.' Similarly, *þurh geswicenysse yfeles* 'by cessation from evil' is literally 'through cessation evil-of', and *cyninges botl* 'a king's palace' is 'king-of palace.' Since the genitive ending already sufficiently performed the function of expressing possession, the preposition *of* did not mark possessive phrases in Old English. Prepositional marking becomes much more frequent in Middle English. The latter is characterized by a common use of the so-called 'group genitive' constructions like *the king of France's son* corresponding to *þæs cyninges sune Frances* 'the-of king-of son France-of' in Old English.

This change was caused by the increasing loss of case inflections and the resulting increased importance of function words.

In conclusion, the possessive markings *'s* and *of* are usually realizations of genitive case in English and establish a grammatical relation between two nouns, but English *'s* also has a lexical use. The manifestation of genitive case varies across languages and historically also within a given language.

References

Allen, Cynthia L. 1997. The origins of the 'Group Genitive' in English. Transactions of the Philological Society 95.1. 111–31.

Anderson, Mona. 1983. Prenominal Genitive NPs. The Linguistic Review 3. 1–24.

Chomsky, Noam. 1970. Remarks on nominalization. Readings in English transformational grammar, ed. by Roderick A. Jacobs and Peter S. Rosenbaum. Waltman, MA: Ginn and Company.

Chomsky, Noam. 1986. Knowledge of language: its nature, origin, and use. New York: Praeger.

McCawley, James D. 1988. The structure of noun phrases. The syntactic phenomena of English, 2 vols. Chicago and London: The University of Chicago Press.

Taylor, John R. 2001. Possessives in English: an exploration in cognitive grammar. New York: Oxford University Press.

EGOR TSEDRYK

See also **Case**

Pragmatics

Pragmatics is the study of language use. Whereas the study of grammar focuses on the language system, pragmatics offers a complementary perspective on language, providing an insight into the linguistic choices that users make in social situations. Pragmatics is, for instance, interested in how people pay compliments, engage in small talk, or write e-mails. The communicative functions of utterances or texts, and the speakers' or writers' intentions behind them, are of particular interest. Historically, the emphasis was on spoken language.

The term 'pragmatics' was first used in semiotics, the general theory of signs. In this theory, pragmatics pertains to the relationship between signs and their users. In linguistics, pragmatics deals with verbal signs (words, utterances, texts) and how they are used by humans in communication. The term 'pragmatics' is derived from Greek *pragma*, which means 'action'. Action is defined as intentional behavior. Pragmatics studies verbal communication as a complex form of intentional behavior, which is interactive, i.e. partner oriented, and symbolic, i.e. conventionalized and culture specific. In the English-speaking world, pragmatics was originally considered part of sociolinguistics.

Pragmatics is a relatively young field of study. So far, there is no coherent pragmatic theory. Pragmatics developed from linguistic, philosophical, and sociological approaches to language use. The so-called pragmatic turn in linguistics was brought about by the writings of a group of philosophers known as speech act theorists in the late 1960s. Until then, mainstream linguistics had focused on linguistic forms and structures, neglecting meaning and ignoring communicative functions and language users.

Initially, pragmatics was dominated by the speech act theory. The basic insight of this tradition is that speech is action. Whenever we say something, we do not only produce sound waves, utter words, and produce sentences but we also perform an action. If a mayor says, *I declare this bridge open*, then this bridge is open. In this sense, the mayor has performed an action.

Speech act theory aims at establishing how many fundamentally different ways there are of doing things with words. Bridge-opening belongs to 'declarations', a class of speech acts that all require an institutional context. Four further classes have been identified, which are more likely to occur in everyday communication. These are 'directives', 'commissives', 'expressives', and 'assertives'. Typical examples for these classes are requests, promises, apologies, and statements, respectively.

In linguistics, the study of language use focuses on the individual speech acts that belong to these fundamental classes, e.g. requests, promises, and apologies, and on the linguistic resources used to perform them. Empirical pragmatics aims at systematically determining the different strategies that people actually use to realize a particular speech act. A classic case is requesting. For instance, if you want a person to take out the garbage, you may say: *Take out the garbage!*, *You ought to take out the garbage*, *How about taking out the garbage?*, *Can you take out the garbage?*, *Garbage day tomorrow*, etc. These strategies differ in their degree of directness. The most direct strategy is to use an imperative construction, and the least direct strategy is to only hint at the desired action. Directness must not be confused with politeness. The degree of directness is a feature of any strategy independent of context, whereas the politeness value of an utterance depends entirely on the social situation.

Theories of verbal politeness have also contributed to the development of pragmatics. According to one particular politeness theory, any speech act may, at least potentially, interfere with the wants and needs of hearers. This is obvious if a speaker wants a hearer to do something. It is less obvious in the case of social niceties, but a compliment, for example, may also be interpreted as harassment. Given the threatening potential of all speech acts, speakers are thought to calculate the imposition involved in the respective situation and then select from a range of direct and indirect strategies available for realizing the act in question. Also, any of these strategies can be modified in a number of ways. For instance, speakers can say, *Could you perhaps take out the garbage, darling?*, where *perhaps* and *darling* serve to modify the basic strategy, *Could you take out the garbage?* Modifications are motivated by polite considerations and aim at reducing the social risks involved in performing the act in question.

A distinction can be made between pragmalinguistics and sociopragmatics. 'Pragmalinguistics' is concerned with the verbal resources available for realizing any given speech act. By contrast, 'sociopragmatics' focuses on the polite norms governing the selection of resources relative to social situations. For example, pragmalinguistics identifies the word choices, meaning patterns, and sentence constructions that are used to pay a compliment, whereas sociopragmatics determines who may compliment whom on what in which situations. Pragmalinguistics is language specific, and sociopragmatics culture specific.

In empirical pragmatics, five types of inquiry can be distinguished. Studies may focus on speech act realization in one language alone, or they may compare speech acts across languages. The second type is known as contrastive pragmatics. It aims at answering questions such as these: how do speakers of English and speakers of Japanese formulate complaints? By contrast, cross-cultural pragmatics examines differences and similarities in social norms and cultural values governing polite choices. A typical question is this: how do members of US American culture and members of British culture refuse offers? A fourth branch of empirical pragmatics is interlanguage pragmatics, which is the study of language use in a second or foreign language. The fundamental question here is the following: how do language learners perform and interpret speech acts in the target language? Finally, applied pragmatics focuses on problems of communication in practical contexts, such as courtrooms, medical interviews, or international business encounters.

Both natural and elicited data are analyzed in empirical pragmatics. Natural data, which occur in everyday situations, are recorded by using field notes, audiotape, or video. Further sources are large corpora of spoken language, such as the Santa Barbara Corpus of Spoken American English. Alternatively (or additionally), a wide range of research instruments is used to elicit data. Tools include role plays, questionnaires, and interviews. Fictional material is also used, e.g. dialogue in narrative prose or drama.

All approaches to language use discussed so far focus on speakers and their intentions, rather than on hearers and their interpretations. How hearers understand utterances is dealt with in a theory of conversational maxims, which also developed in speech act philosophy. According to this theory, hearers assume that speakers behave cooperatively at all times (cooperative principle). More specifically, hearers expect speakers to observe a set of conversational maxims, involving 'Don't lie' (maxim of quality), 'Say as much as is necessary, no more and no less' (maxim of quantity), 'Be relevant' (maxim of relation), and 'Be brief' (maxim of manner). If an utterance does not, for example, seem to relate to

what was said before, or if maybe more or less is said than is necessary, then the hearer tries to work out an interpretation that makes sense in the given context. This process is called 'inferencing'. A pragmatic theory of communication and cognition that focuses on the process of understanding utterances is relevance theory. It is based on the conversational maxim of relation.

In everyday communication, people do not always observe the maxims of conversation. For example, people lie, talk too much, or make obscure statements. Such behavior can be explained by assuming that a politeness principle may conflict with the cooperative principle. In a particular situation, conversationalists may rate politeness maxims higher than the conversational maxims. For instance, a 'white lie', which violates the maxim of quality, can sometimes be more considerate than telling the truth. Generally speaking, maxims of politeness minimize unfavorable effects for the hearer or increase favorable effects. For example, positive evaluations expressed in compliments are often maximized, whereas impositions involved in requests are usually reduced.

Most theories of language use and most empirical approaches are speech act-based, i.e. they focus on utterances in isolation. Speech act-based pragmatics is referred to as micropragmatics. Its shortcomings are remedied in macropragmatics, which provides a fuller picture of language use by accounting for the dialogical and interactive nature of human communication. Macropragmatic approaches do not focus on the speaker or hearer alone, but rather on participants in communication who alternately adopt the role of speaker or hearer. Also, they challenge the view that the speech act is the central communicative unit and claim that the 'exchange' is more important. An exchange minimally consists of two speech acts. The first act functions as an initiating 'move', and the second act as a responding 'move'. Moves indicate the interactive function of an act relative to the immediate verbal context. Typical exchanges are question–answer, offer–refusal, and greeting–greeting. Between initiation and response, countering moves can occur, as in *Can you take out the garbage?* (initiating move) *Why don't you take it out yourself?* (countering move) *Okay, okay. Calm down, I will* (responding move). An exchange is closed when an outcome has been reached.

In macropragmatics, complete communicative events such as conversations, interviews, debates, etc. are analyzed. Communicative events are considered hierarchical structures. The hierarchy, starting with the smallest unit, involves acts, moves, exchanges, sequences, and phases. At each of these levels, there are obligatory units (heads) and optional units (supportives). Thus, we distinguish head moves and sup-

portive moves, head exchanges and supportive exchanges, and so on. For example, the utterance *Can you lend me ten dollars? I've left my purse at home* consists of a head move, which is the request, and a supportive move, which is an explanation.

The type of macropragmatics sketched here is based on British discourse analysis, which was developed in linguistics. An alternative approach is conversation analysis, which originates in American sociology. In conversation analysis, the term 'adjacency pair' is used for the exchange unit. Another unit identified in conversation analysis is the 'turn'. Conversationalists take turns at talk. A turn is defined as a speaker's contribution at a particular point in a conversation. It is surrounded by speaker switches. Turns may differ considerably in length. Prototypically, a turn consists of a second pair-part to match the previous speaker's first pair-part and a first pair-part that must be matched by the next speaker. In conversation analysis, the mechanisms of turn-taking and turn-allocation for encounters involving two or more speakers have been identified, and simultaneous speech and interruptions have also been investigated. Conversation analysis differs from discourse analysis in using only naturally occurring conversation, whereas discourse analysis also works with other data types, including fictional dialogue.

Most research in pragmatics deals with present-day language. Recently, however, an interest has also developed in earlier periods. Two branches have evolved, namely, historical pragmatics and diachronic pragmatics. Historical pragmatics deals with previous stages of language, e.g. Old English or Middle English. By contrast, diachronic pragmatics studies development across time. In both types, two alternative perspectives can be adopted. On the one hand, the way in which a particular communicative function, e.g. a speech act, was realized at different times may be examined. On the other the researcher may start with a particular expression, e.g. a routine formula, and describe its uses in certain periods or in the course of history. It is not always possible to keep these two perspectives apart. Typical research topics in historical pragmatics include greetings in the fifteenth century and politeness strategies in Shakespeare's plays. Topics for diachronic pragmatics include, for instance, insults in Old, Middle, and Modern English or changes in apologizing through the ages. Although all historical sources are in the written mode, some include a considerable amount of information on spoken language. For example, court reports may incorporate detailed accounts of what defendants or witnesses actually said. Finally, recent and ongoing pragmatic change may be studied in e-mails and chat-room conversations, because netiquette is only just emerging.

Another recent development in pragmatics is the study of synchronic variation in contemporary language use. Sex differences have been examined to some extent, but regional, social, and ethnic variation has not received much attention in pragmatics. Sociolinguistic research suggests that we can expect speech act realizations, politeness markers, discourse strategies, and turn-taking, etc., to vary across regions, urban and rural communities, social classes, and ethnic groups. This is an area for future investigation.

In short, pragmatics is the study of language use. Pragmatics is not a component of the language system, but it offers a different perspective on verbal phenomena. It examines how linguistic resources are used in communication and investigates a speaker's (or writer's) intentions and a hearer's (or reader's) interpretations. Language use is considered a complex form of social action. The general question addressed in pragmatics is how language functions in the lives of human beings. The focus can be on utterances (micropragmatics) or on longer stretches of discourse (macropragmatics). Pragmalinguistics is concerned with the structural resources that a language provides for conveying particular intentions, whereas sociopragmatics considers language use relative to social situations. There is no coherent pragmatic theory to date. Theoretical pragmatics draws on a number of disciplines, including philosophy, sociology, and anthropology. Speech act theory, politeness theory, discourse analysis, and conversation analysis provide major contributions. Empirical pragmatics involves the investigation of speech act realizations, politeness phenomena, and discourse strategies on the basis of natural, elicited, and also fictional data. The comparison of language use across languages and cultures is also an important empirical research component. Applied pragmatics focuses on problems of communication in practical contexts. Most pragmatic research investigates present-day language, but recent developments also include historical and diachronic perspectives.

References

Blakemore, Diane. 1992. Understanding utterances: an introduction to pragmatics. Oxford, Cambridge, MA: Blackwell.

Blum-Kulka, Shoshana, Juliane House, and Gabriele Kasper, (eds.) 1989. Cross-cultural pragmatics. Norwood, NJ: Ablex.

Jucker, Andreas, (ed.) 1995. Historical pragmatics. Amsterdam: Benjamins.

Leech, Geoffrey. 1983. Principles of pragmatics. London and New York: Longman.

Levinson, Stephen. 1983. Pragmatics. Cambridge: Cambridge University Press.

Mey, Jacob (ed.) 1998. Concise encyclopedia of pragmatics. Oxford: Elsevier.

Mey, Jacob. 2001. Pragmatics: an introduction, 2nd edition. Oxford and Cambridge, MA: Blackwell.

Trosborg, Anna. 1995. Interlanguage pragmatics. Berlin and New York: Mouton de Gruyter.

Verschueren, Jef. 1999. Understanding pragmatics. London and New York: Arnold.

Wierzbicka, Anna. 1991. Cross-cultural pragmatics. Berlin and New York: Mouton de Gruyter.

KLAUS P. SCHNEIDER

See also **Speech Acts**

Predication

Although predication is extremely important in contemporary linguistic research, it is a very old concept tracing back to Aristotelian logic, where it emerged to indicate the semantic relationship existing between the event described in a proposition and the individuals participating in such an event. Due to its origin in the Western philosophic tradition, the linguistic analysis of predication has been approached for a long time from a semantic perspective. From the 1980s onward, however, its study took a different course due to the advent of several transformational theories mostly devoted to examine the syntactic behavior of predication.

Being traditionally accounted for in semantic terms as the connection between a predicate—the event expressed in the clause—and its arguments—the entities involved in the event—the semantic analysis of predication has been mainly interested in how this linking takes place. The most common answer to this question is found in Theta Theory—that is, the lexico-semantic module of the transformational theoretical framework known as Government and Binding. Since it is chiefly concerned with the predicate's assignment of theta roles to its different arguments, it specifies the exact number of arguments a particular predicate requires and their semantic nature; that is, if they take part in the predication as agents, patients, experiencers, themes, etc. It must be noted, however, that, being unlimited in number, there is no agreement as regards

the name and catalogue of these detailed semantic traits alloted by a predicate to its arguments, generally called theta roles. Regardless, the main principles underlying Theta Theory are formulated in the well-known Theta-Criterion, which postulates the following restrictions on the assignment of theta roles: (a) every argument in a predication must receive one theta role from its predicate; (b) the same theta role cannot be assigned to more than one participant in the predication. According to Theta Theory, then, and as illustrated in *Mary gave some money to her little brother*, a predicate like *give* needs three arguments bearing the following theta roles to be completed: an agent (*Mary*), a theme (*some money*), and a recipient (*her little brother*).

From the syntactic viewpoint, predication is defined via saturation: a predicate is a maximal projection—that is, a full phrasal category—and, consequently, an open function, which must be syntactically closed by the appropriate arguments to form a grammatically acceptable chain. Despite being the maximal projection with the strongest predicative capacity, the verbal phrase is not the only one able to function as predicate: *The girl read her cousin a book*. The adjectival (*Bill is fond of Jan*), the nominal (*This ratty piece of leather is a wallet*), the prepositional (*Those sisters are on drugs*), and the adverbial phrases (*She feels a bit down today*) can also perform, as illustrated between brackets, the predicative function. Although different predicates demand a different number of arguments, as the previous examples show, they all share the property of having, at least, one compulsory argument: their subject. This argument (*The girl, Bill, This ratty piece of leather, Those sisters* and *She* in the previous examples) differs from the rest of possible arguments that a predicate may have both semantically and syntactically: semantically, because it constitutes the starting point for the predication; that is, it is the participant the predication is about; and syntactically, because it is the predicate's only external argument. Although the subject and predicate constituents are unanimously said to be in a relation of mutual c-command—that is, they are syntactic sisters dominated by exactly the same maximal projections—the two main transformational approaches to the study of predication explain the subject–predicate syntactic relationship in different ways: the Predication Theory, for instance, considers that subject and predicate are independent structural components, whose syntactic relation has to be marked by indexing at surface structure, and the Clausal Theory defends, in turn, that they both form a single constituent having clausal status either with or without verb. If this is the case, the clause containing the predication relationship is catalogued as small. An example of a small clause is, thus, represented by *Jill a nice person* in *Rachel considers Jill a nice person*.

Although the previous syntactico-semantic characterization applies to predication in general, some refinements must be made because not all instances of predication exhibit the same grammatical behavior. Predication cannot be considered, therefore, a homogeneous linguistic phenomenon, but rather a heterogenous one involving two distinct classes of predication: primary and secondary predication. The former corresponds to the prototypical notion of predication because, being syntactic and semantically independent, it embodies the basic and matrix predicative relationship occurring at the level of the clause: *John painted the house, John eats carrots*. The latter lacks, in turn, such syntactico-semantic autonomy and must always appear, consequently, complementing an instance of primary predication: *John painted the house red, John eats the carrots raw*. The aforementioned contrast is, furthermore, extremely important because it determines the status of the subject of each type of predication: whereas the subject of a primary predication is systematically, like *John* in the previous examples, the predicate's theta-marked external argument, the subject of a secondary predication displays a dual function: it is the secondary predicate's theta-marked external argument at the same time as a theta-marked argument of the primary predication; notice, for instance, that in the examples above, *the house* and *the carrots* are, on the one hand, the subjects of the secondary predicates *red* and *raw*, and on the other, the direct objects of the primary predicates *painted* and *eats*.

Besides these contrasts, two semantic types of secondary predication must be distinguished: resultative and depictive secondary predication. The former, as illustrated by *the house red* in the previous example, signals a potential result of the action denoted by the verb and can only be predicated by the direct object argument of the primary predication; the latter, in turn, does not maintain any semantic relation with the verbal predicate it complements, because, as shown by *raw* in the aforementioned example, depictive secondary predicates denote intrinsic and temporary properties of the subject or direct object arguments of the primary predication, which must exist at the time of the verbal action.

References

Aarts, Bas. 1995. Secondary predicates in English. The verb in contemporary English. Theory and practice, ed. by Bas Aarts and Charles F. Meyer. Cambridge: University Press.

Dik, Simon C. 1981. Predication and expression: the problem and the theoretical framework. Predication and expression in functional grammar, ed. by A. Machtelt Bolkestein, Henk A. Combé, Simon C. Dik, Casper de Groot, Jadranka Gvozdanovi_, Albert Rijksbaron, and Co Vet. London: Academic Press.

Frawley, William. 1992. Linguistic semantics. New Jersey: Lawrence Erlbaum Associates.

Nakajima, Heizo. 1990. Secondary predication. The Linguistic Review 7. 225–309.

Napoli, Donna Jo. 1989. Predication theory. A case study for indexing theory. Cambridge: Cambridge University Press.

Rapoport, Tova R. 1990. Secondary predication and the lexical representation of verbs. Machine Translation 5. 31–55.

Rothstein, Susan D. 1985. The syntactic forms of predication. Bloomington: Indiana University Linguistics Club.

Stowell, Tim. 1983. Subject across categories. The Linguistic Review 2. 285–312.

Williams, Edwin S. 1980. Predication. Linguistic Inquiry 11. 1. 203–38.

Williams, Edwin S. 1983. Against small clauses. Linguistic Inquiry 14. 2. 237–308.

BEATRIZ RODRÍGUEZ-ARRIZABALAGA

See also **Thematic Structure**

Professions for Linguists

Linguists study aspects of language, languages, and language use. With this diversity of focus, linguists may be found in many different professions. Linguists are used in both private and public sectors, in jobs related directly to linguistics, as well as in professions where a knowledge of language or language use is highly relevant, although not directly required.

The nature of language has interested scholars for over 2,000 years. The earliest works of a linguistic nature date to some of the first writings known to humankind. Such early linguistic writings include discourse on the nature of language and humanity (e.g. Socrates), the reconstruction of the spoken language of the Vedas (e.g. Panini in India), and the translation of texts and codification of languages (e.g. the trilingual Rosetta Stone, Varro's grammar of Latin). The description of child language acquisition has also fascinated scholars through the ages, including scientists (e.g. Charles Darwin), psychologists (e.g. Piaget), philosophers (e.g. Rousseau), and others. In fact, the disciplines within which linguists have worked have included philosopher (e.g. Plato, Locke, Hume), grammarian (e.g. Dionysius Thrax, Cicero), philologist (e.g. Saussure), anthropologist and field linguist (e.g. Boas, Sapir), psychologist (e.g. Skinner, Piaget), scientist (e.g. Bloomfield), and 'linguist' within an independent field of research (e.g. Chomsky).

In their professions, linguists may focus on the sounds of the language (e.g. phonetics, phonology), the words of the language (e.g. semantics), how words combine (e.g. syntax, discourse), how speakers acquire and/or use language (e.g. sociolinguistics, psycholinguistics, applied linguistics, dialectology), how the human brain processes language (e.g. psycholinguistics, neurolinguistics), how to program machines to carry out voice recognition/voice synthesis (e.g. computational linguistics), or even how animal communication systems, or languages, work (e.g. zoosemiotics).

The sounds of language are the focus of phoneticians in the fields of acoustic phonetics (e.g. the study of the physical properties of speech sounds using instrumental techniques of investigation to provide an objective account of speech patterns, which can be related to the way sounds are produced and heard), auditory phonetics (e.g. the study of the way people perceive sound, as mediated by the ear, auditory nerve, and brain), or articulatory phonetics (e.g. the study of the use of the vocal organs (articulators) to produce the sounds of speech).

The sound systems of language are of interest to phonologists, who study the sound systems of languages and the general or universal properties displayed by these systems.

Phonetics and phonology are also relevant to the teaching of pronunciation, speech pathology and speech therapy, accent reduction, voice and drama coaching, elocution and public speaking, forensic linguistics (e.g. voice recognition, voice printing), sound engineering, and the growing area of speech synthesis and voice activation in the computer industry.

Syntacticians study the structure (or grammar) of languages, including the way words are combined to form sentences, relationships between the elements of sentence structure, and the rules governing the arrangement of sentences in sequences. Meaning in language, in terms of relations such as synonymy and antonymy, is the domain of semanticians. Issues such as to what extent semantic concepts are universal is a current area of focus.

In addition to academic research, syntacticians and semanticists work on dictionaries and grammars of different languages, develop and study computer languages and computer translation, preserve endangered and dying languages through field research, provide forensic analysis for the legal system, develop language tests, develop language policy and other educational curriculum, work in language reconstruction

(with archeologists), etc. Linguists also work for intelligence agencies as spies, translators, interpreters, code-breakers, and computer programmers.

Professions for linguists related to the acquisition of first/child language include developmental linguistics and developmental psychology, psycholinguistics, early childhood education, deaf education, teaching, speech therapy, speech pathology, textbook and curriculum development, and software development. Similarly, professions for linguists related to the acquisition of second or additional languages include applied linguistic research, ESL/EFL teaching, foreign and modern language teaching, bilingual/bicultural education, teacher training, translation, proofreading, marketing, textbook writing and publishing, editing, speech therapy, speech pathology, software development, website design and marketing, on-line teaching, recruiting students for language schools (e.g. agents), owning language schools and other related businesses, developing and implementing multicultural policies, facilitating minority language rights, language policy, language planning in business and government services, and as freelance consultants (e.g. providing program development/analysis, giving seminars on gender issues in the workplace, cross-cultural training, etc.).

Some additional areas within linguistics in which linguists are employed include anthropological linguistics (the study of languages in relation to social and cultural patterns and beliefs); areal linguistics (the study of geographical areas that are characterized by shared linguistic properties); audiology (the study of hearing and hearing disorders) biolinguistics (the study of the biological preconditions for language development and use); dialectology (the study of dialects, dialect geography, or linguistic geography); *historical linguistics* (the study of language variation and/or change); *neurolinguistics* (the study of the basis in the human nervous system for language development and use); philology (the study of language history, including historical study of literary texts); *philosophical linguistics* (the study of the role of language in relation to philosophical concepts, as well as the philosophical status of linguistic theories, meth-

ods, and observations); pragmatics (the study of the conditions on language use deriving from the social situation); semiology (the study of signs and their use, focusing on the mechanisms and patterns of human communication and on the nature and acquisition of knowledge); semiotics (the study of language as one type of sign system, along with bodily gestures, clothing, and the arts); sociolinguistics (the study of how language is integrated with human society with reference to race, ethnicity, class, sex, and social institutions); and zoosemiotics (the study of animal communication systems).

In the new millennium, with globalization and the explosion of new communication technologies, the knowledge and expertise of linguists are increasingly in demand in the fields of education, business, technology, health, law, and communications. The field that began with philosophical explorations of language and humanity is now recognized as being in the forefront of human/machine interfaces in the twenty-first century.

References

Brand, Amy. A short road from academic research to academic publishing: one linguist's trip. Linguistic Enterprises Advice. http://web.gc.cuny.edu/linguistics/enter/elgin.htm.

Crystal, David. 1987. The Cambridge encyclopedia of language. Cambridge, UK: Cambridge University Press.

Haden Elgin, Suzette. Another item in the freelance linguist's survival kit: Seminars, Linguistic Enterprises Advice, http://web.gc.cuny.edu/linguistics/enter/elgin.htm

Levi, Judith, Questions and answers about language consulting with lawyers, Linguistic Enterprises Advice, http://web.gc.cuny.edu/linguistics/enter/elgin.htm

O'Grady, William, and Michael Dobrovolsky. 1996. Contemporary linguistic analysis. Toronto, CA: Copp Clark Ltd.

Wylie, Dovie. But Mom, who is going to pay you to know these things? Linguistic Enterprises Advice. http://web.gc.cuny.edu/linguistics/enter/elgin.htm

Other Work-Related Websites

Dave's ESL Café: http://www.eslcafe.com/search/jobs/

Linguist List: http://www.emich.edu/~linguist/jobsindex.html

American Association of Applied Linguistics: http://www.aaal.org

Linguistic Enterprises: http://web.gc.cuny.edu/linguistics/enter/

KAREN WOODMAN

Proficiency Testing

The term proficiency *testing* embodies two main viewpoints: an earlier, traditional one referring to tests of ability to use a language for some extralinguistic purpose, and the more recent 'Proficiency Movement'

in foreign language assessment associated with the American Council on the Teaching of Foreign Languages (ACTFL).

Traditionally, proficiency testing has been contrasted mainly with achievement testing in which the test content and method are based on a language teaching syllabus. The purpose of an achievement test is to find out how much students have learned of what has been taught. By contrast, proficiency tests are not linked to any syllabus but rather attempt to measure how well a learner can use the language regardless of the circumstances in which it was acquired. Lado (1961) was the first to make this contrast explicit:

> [Achievement] tests...attempt to measure how much of a language a student knows. When the...tests are thought of independently of the learning experience, they may then be referred to as proficiency tests. Proficiency tests measure how much of a foreign language a person (not necessarily a student) knows. (p. 369)

Note that Lado speaks of the purpose of the proficiency test as measuring how much of the language a person knows, reflecting his 'discrete point' approach to test development wherein specific grammatical points are tested more or less in isolation. At about the same time, other scholars were calling for proficiency tests that required language performance in contexts of use. For example, Carroll (1961 [1972]) argued that

> an ideal English language proficiency test should make it possible to differentiate, to the greatest possible extent, levels of performance in those dimensions of performance which are relevant to the kinds of situations in which the examinees will find themselves after being selected on the basis of the test. (p. 319)

Thus, Carroll explicitly links proficiency testing to performances relevant to future situations, and this has become the standard view of proficiency: ability to use a language for some extralinguistic purpose (Davies et al. 1999).

The extralinguistic purpose that proficiency testing is most often associated with is admission to academic programs. For example, the two best-known proficiency tests, the Test of English as a Foreign Language (TOEFL) and the International English Language Testing System (IELTS), are both used primarily for admission to university programs in the United States, the United Kingdom, and Australia. These tests were developed to answer the question of whether a candidate for admission to an English-speaking academic institution possesses sufficient English ability to cope with academic work.

However, proficiency tests need not be limited to academic situations only. The focus on future uses of the language naturally leads to tests of languages for specific purposes (LSP). For example, the *Occupational English Test* (OET) was developed in Australia to ascertain whether migrant candidates for licensure in the health professions are able to use English with sufficient ability to work with English-speaking patients and clients; the Japanese Language Test for Tour Guides was also developed in Australia; the Test of English for International Communication (TOEIC) is a US test aimed primarily at international business people; and the Proficiency Test in English for Air Traffic Controllers is a European test for trainee air traffic control officers. These and other specific-purpose language tests were developed without reference to any particular course of instruction but rather on the basis of an analysis of a particular target language use situation (Douglas 2000).

The distinction between achievement and proficiency testing is not entirely rigid, either. For example, Bachman (1990) argues as follows:

> Whether or not the specific abilities measured by a given proficiency test actually differ from those measured by a given achievement test will depend, of course, on the extent to which the theory upon which the proficiency test is based differs from that upon which the syllabus is based (p. 71).

Note that Bachman defines proficiency tests as theory-based, and indeed, it is generally the case that proficiency tests are related to some theory of what it means to know and use a language. Since Hymes's (1972) formulation of communicative competence, proficiency test developers have used theories about the components of language knowledge to guide them in deciding what to test. The current, most well-known framework is that proposed by Bachman (1990) and Bachman and Palmer (1996) of communicative language ability, consisting of grammatical, textual, functional, and sociolinguistic language knowledge, plus strategic competence, which directs the use of one's language knowledge in actual communication.

The so-called Proficiency Movement, although more recent than proficiency testing itself, has its origins in the early days of the assessment of the oral language skills of US Government employees, most notably the Foreign Service Institute (FSI), the Peace Corps, the Central Intelligence Agency (CIA), and the military (Lowe 1988). The FSI (now the Interagency Language Roundtable [ILR]) Oral Proficiency Interview became the standard for the testing of spoken language ability, and the American Council on the Teaching of Foreign Languages referred to it in responding to a mandate from the 1979 President's Commission on Foreign Languages and International Studies to establish a series of nationally recognized descriptors that would facilitate assessment of the proficiency of both students and teachers in foreign language programs. The ACTFL Proficiency Guidelines

(1986, 1999) thus 'represent the combined efforts of groups of educators to provide an operational definition for [proficiency], to represent it as phenomena observable and evolving' (Galloway 1987: 26). The Guidelines are directed primarily at assessing the foreign language proficiency of college and university students in the United States, and consist of descriptors of performance in speaking, listening, reading, and writing at Superior, Advanced, Intermediate, and Novice levels, with the latter three levels divided into high, mid, and low categories.

The focus of the Proficiency Movement, as embodied in the ACTFL Guidelines, is the assessment of language production or performance rather than internalized language knowledge or ability for use (Galloway 1987; Lowe 1988), thus distinguishing the Movement from the more traditional understanding of the concept of proficiency described above.

References

American Council on the Teaching of Foreign Languages. 1986. ACTFL proficiency guidelines. Hastings-on-Hudson, NY: American Council on the Teaching of Foreign Languages.

American Council on the Teaching of Foreign Languages. 1999. ACTFL proficiency guidelines—Speaking. Hastings-on-Hudson, NY: American Council on the Teaching of Foreign Languages.

Bachman, Lyle. 1990. Fundamental considerations in language testing. Oxford: Oxford University Press.

Bachman, Lyle, and Adrian Palmer. 1996. Language testing in practice. Oxford: Oxford University Press.

Carroll, John B. 1972. Fundamental considerations in testing for English language proficiency of foreign students. Teaching English as a second language: a book of readings, ed. by H.B. Allen and R.N. Campbell. New York: McGraw-Hill. (Originally published in Testing, Washington, DC: Center for Applied Linguistics, 1961.)

Davies, Alan, Annie Brown, Cathie Elder, Kathryn Hill, Tom Lumley, and Tim McNamara. 1999. Dictionary of language testing. Studies in language testing, Vol. 7. Cambridge: Cambridge University Press.

Douglas, Dan. 2000. Assessing languages for specific purposes. Cambridge: Cambridge University Press.

Galloway, Vickie. 1987. From defining to developing proficiency: a look at the decisions. Defining and developing proficiency: guidelines, implications, and concepts, ed. by Heidi Byrnes and Michael Canale. Lincolnwood, IL: National Textbook Company.

Hymes, Dell. 1972. On communicative competence. Sociolinguistics, ed. by J.B. Pride and J. Holmes. Harmondsworth, UK: Penguin.

Lado, Robert. 1961. Language testing: the construction and use of foreign language tests. London: Longman, Green, and Company.

Lowe, Pardee. 1988. The unassimilated history. Second language proficiency assessment: current issues, ed. by Pardee Lowe and Charles Stansfield. Englewood Cliffs, NJ: Prentice-Hall Regents.

McNamara, Tim. 1996. Measuring second language performance. London, New York: Longman.

Dan Douglas

See also **Second Language: Learning; Second Language: Teaching**

Pro-forms

In linguistics, a pro-form is generally taken as an element used in place of other linguistic element(s). Pronouns are the most common pro-forms used to substitute for a noun or a noun phrase. 'She', for example, is a third-person singular pronoun used in place of a singular human female animate noun such as a woman as in, 'A woman is coming to see you. *She* has called you earlier.'

Although pronouns have often been used as examples of pro-forms, there are other linguistic elements that have comparable properties but do not substitute for a noun or a noun phrase. '*So*', as in 'He thinks it will rain tonight but I don't think *so*,' substitutes for a whole clause, i.e. 'it will rain'.

Here and *there* are pro-forms for locative expressions; bare noun phrases may be substituted with *what* and *one/ones*, and adjectives by *such* and verb phrases by *do*.

The term 'pro-form' was probably first used by Jerrold Katz and Paul Postal (1964) as a mechanism to explain both syntactic and semantic aspects of the substitutions in the above examples. Syntactically, the pro-constituent guarantees the recoverability of a substitution or deletion. Semantically, the pro-form calls for interpretation by retrieving its equivalents.

The term pro-form, since its introduction, has often been used alternately with pronoun, and now it seems to replace pronoun. However, some linguistic elements seem to have comparable properties to pronouns but they are not substitutes for nouns. In fact, there are many other classes of words than nouns that get a different form in the following mention in a text.

Thus, the term pro-form seems appropriate to be used collectively for any kind of substitution.

In the literature of generative grammars, a pro-form is often found as an element that assumes the process of substitution. In these theories, pro-forms can be used as one of the key tests for constituency in syntax: if a string of words can be replaced by a pro-form, these words form a constituent, i.e. a structural unit within the sentence. To illustrate, in the sentence 'The student read a book in the library,' the pronoun *she* can substitute for the noun phrase 'the student', a *there* can stand in for the prepositional phrase 'in the library', and a *did* can substitute for the verb phrase 'read a book in the library'. Consequently, we may conclude that these three types of phrases are constituents.

Moreover, there are some other terms that are loosely used in place of pro-forms. One of these is ellipsis. Andrew Radford (1997) considers ellipsis a process by which redundant information in a sentence is omitted. Pro-forms, however, are not omissions but substitutions. However, some recent work in natural language processing (e.g. by Daniel Hardt 1993) includes pro-form as one category of elliptical forms.

In terms of semantics, a pro-form has no meaning in itself, rather it requires a retrieval of meaning from a previously mentioned element, or antecedent, i.e. the element for which it substitutes. In other words, pro-forms are semantically bound by other elements.

References

Hardt, Daniel. 1993. Verb phrase ellipsis: form, meaning and processing. Ph.D. dissertation. University of Pensylvania.

Katz, Jerrold, and Paul Postal. 1964. An integrated theory of linguistic descriptions. Cambridge: Cambridge University Press.

Radford, Andrew 1997. Syntactic theory and the structure of English: a minimalist approach. Cambridge: Cambridge University Press.

Van Valin, Robert D. Jr. 2001. Introduction to syntax. Cambridge: Cambridge University Press.

PRANEE KULLAVANIJAYA

Proper Nouns

Proper nouns (usually known as 'proper names' by philosophers) are basically names of specific people, places, organizations, months, festivals, and so forth. Typical examples are *John*, *London*, *Oxfam*, *January*, and *Christmas*. The investigation of the semantic meaning of proper nouns goes back to the Greeks. In 200 BCE, the Greek grammarian Dionysius Thrax gave one of the most often used definitions: a proper noun signifies an individual being, whereas a common noun signifies a general substance. As time has gone by, philosophers (e.g. John Stuart Mill, Gottlob Frege, Bertrand Russell, and Saul Kripke) and linguists (e.g. Otto Jespersen, Alan Gardiner, and John Algeo) have put forward a number of criteria for the semantic description of proper nouns with reference to the traditional assumption of a proper–common dichotomy.

One of the earliest and most famous criteria was introduced by the English philosopher John Stuart Mill in *A system of logic* (1843). He suggests that a pure proper noun is a denotative sign without connotation. In contrast to a common noun that both denotes and implies attributes, a proper noun is used to denote a specific or unique individual entity regardless of what properties it has. For example, a proper noun like *Victoria* simply denotes the individuals who are called by *Victoria*. It does not indicate any attributes as belonging to these individuals. Sometimes, we do not even know whether it is a woman's name or a place name. With reference to the semantic information about the entities, pure proper nouns are lexically opaque and the entities they designate are normally unpredictable. On the other hand, a common noun is used to subsume similar entities under a generic concept. For example, various books, irrespective of material, size, or purpose, are subsumed under the generic concept *book*. Thus, the use of a common noun for an entity is completely determined by the meaning of the noun. Common nouns such as *chair* are semantically transparent because they have lexical meaning and carry information.

It has been argued that while proper nouns have no meaning in isolation, they carry descriptive meaning when applied in a specific context to a particular person or place. For example, the proper noun *John* is used in English-speaking countries as a name for a male. However, this culturally specific information should not be identified with the meaning carried by a common noun. The information associated with a proper noun is usually changeable, whereas the meaning of a common noun is basically constant.

Within a given context of utterance, proper nouns together with pronouns are two distinct kinds of singular definite referring expressions that enable the hearer to denote the actual referent from the class of potential referents. Unlike pronouns that have some descriptive meaning, proper nouns have reference but not sense. They identify their referents by making use of the unique and arbitrary association that holds between a proper noun and its referent. Moreover, unlike pronouns that are used by different speakers to denote different entities in different contexts, proper nouns are always claimed in any particular context to refer to one and only one entity, although a proper noun may name many individuals. It follows that a proper noun merely serves as an identification mark by singling out an entity from among similar entities. Its function is precisely that of labeling.

In the 1970s, the American philosopher Saul Kripke (1972, 1980) espoused the view that proper nouns are 'rigid designators' that designate the same individual in all possible states of the world whatever their characteristics happen to be. According to Kripke, the proper name *John* is a rigid designator because it picks out particular individuals who happen to be *John* without any defining characteristics in whatever state of affairs we entertain. However, the common noun *chair* refers to the same thing in the world that happens to fit the defining characteristics.

In some languages such as English, it has long been held that the definition for typical proper nouns should also be laid down on the basis of three additional criteria: orthographic, morphological, and syntactic.

Orthographically, it is a graphic convention that English proper nouns are marked with initial capitalization. By the morphological criterion (i.e. the relationships between the parts that make up the proper noun itself when it is more than a single morpheme), a typical English proper noun has no plural suffix in the strictest sense e.g. *Britain—*Britains*. By the syntactic criteria (i.e. the relationships of the proper noun with other items in the same grammatical construction), typical English proper nouns are not preceded by the article or other determiners since the references that proper nouns have is by their very nature self-determining or inherently definite. Unlike common nouns, proper nouns normally lack the contrastive definiteness (e.g. **an America*, **some Americas*; *an orange*, *some oranges*). Another major syntactic criterion is that proper nouns lack modification. When they have the normal unique denotation, they can only accept nonrestrictive modifiers (e.g. *Mary, who is my sister, is studying abroad—*Mary who is my sister is studying abroad.*). Proper nouns require *the* when they are modified by a restrictive clause (e.g. *The Mary I know is my friend's wife*), but not when they are unrestricted.

While numerous philosophers and linguists have considered proper nouns and common nouns as two discrete categories and the boundary between them as unambiguous, Jespersen (1924) holds that no sharp line can be drawn between proper and common nouns, the difference being one of degree rather than of kind. Later, Quirk et al. (1985) explicitly stated that the class of proper nouns has unclear boundaries and the degree of membership involves the notion of gradience. There are some circumstances in which proper nouns (e.g. *the Avon*, *the Crimea*, *the Himalayas*) behave like common nouns by taking the definite article and the plural form. On the other hand, a number of common nouns with unique denotation are sometimes capitalized and thus enjoy a similar semantic function as proper nouns e.g. *Fortune*, *Wealth*, *Fate*.

References

Algeo, John. 1973. On defining the proper name. Gainesville, FL: University of Florida Press.

Frege, Gottlob. 1892. Über Sinn und Bedeutung. Zeitschrift für Philosophie und philosophische Kritik, Vol. 100; as On sense and reference, translated by Max Black, in Translations from the Philosophical Writings of Gottlob Frege, ed. by Peter Geach and Max Black. Oxford: Basil Blackwell, 1960.

Gardiner, Alan. 1940. The theory of proper names: a controversy essay. London: Oxford University Press; 2nd edition, 1954.

Jespersen, Otto. 1924. The philosophy of grammar, London: Allen and Unwin; New York: Norton, 1965; Chicago and London: The University of Chicago Press, 1992.

Kripke, Saul A. 1972. Naming and necessity. Semantics of natural language, ed. by G. Harman and D. Davidson. Dordrecht, Netherlands and Boston, USA: D. Reidel Publishing Co.; as Naming and necessity. Oxford: Basil Blackwell, revised edition, 1980.

Lyons, John. 1977. Semantics, 2 vols. Cambridge: Cambridge University Press.

Marmaridou, Sophia A.S. 1991. What's so proper about names? A study in categorisation and cognitive semantics. Athens: University of Athens.

Mill, John Stuart. 1843. A system of logic, ratiocinative and inductive, being a connected view of the principles of evidence, and the methods of scientific investigation. London: John W. Parker; in Collected works of John Stuart Mill, Vol. VII, ed. by J.M. Robson. Toronto and Buffalo: University of Toronto Press, and London: Routledge and Kegan Paul, 1973.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. A comprehensive grammar of the English language. London and New York: Longman.

Russell, Bertrand. 1950. An inquiry into meaning and truth. London: George Allen and Unwin; London and New York: Routledge, revised edition, 1995.

Ullmann, Stephen. 1962. Semantics: an introduction to the science of meaning. Oxford: Basil Blackwell.

Zabeeh, Farhang. 1968. What is in a name? An inquiry into the semantics and pragmatics of proper names. The Hague: Marinus Nijhoff.

GRACE YUEN WAH TSE

See also **Naming; Reference**

Prosody

Prosody refers to the (variations in) rhythm, intonation, and stress patterns in speech. Its role is to emphasize words, to segment a sentence into smaller units, to change the meaning of an utterance (not of a speech sound!), and to resolve syntactic disambiguity (e.g. 'They fed her dog biscuits' Who ate the biscuits? She or the dog?). Apart from these language-dependent aspects, prosody is also used in a more universal manner, to provide cues to the state of the speaker and to convey the speaker's mood (happiness, impatience, boredom, etc.). Prosodic patterns are very important during speech development. They appear in the infant's speech long before the first word: babies often speak nonsense with the correct intonation! Moreover, as boundaries between words are not clearly marked in speaking, infants use prosodic information to 'segment out' words in the continuous stream of speech sounds. This process is termed 'prosodic boot-strapping'. And the importance of prosody also becomes clear when listening to foreign speakers: even if they articulate well, they are not easily understood if their use of stress and timing is incorrect. The same can be said of deaf speakers: their speech is often unintelligible because they cannot master the prosody of their language well. In noisy environments prosody helps the listener to understand the speaker. Moreover, spectrally distorted speech with the original prosody is easier to understand than the original speech signal with distorted prosody. Prosody does not only contribute significantly to the intelligibility of speech but also to its naturalness. Although present-day speech synthesis is highly intelligible, it still sounds artificial. More research on speech prosody is needed to make it sound natural or pleasant enough to listen to for long stretches of time.

The most important prosodic features are intonation, stress, and quantity. The acoustic correlates of these features are fundamental frequency (perceived as pitch), intensity, and duration (although there is no one-to-one correspondence).

Intonation

Variations in tone (word level) and intonation (sentence level) are used to change the meaning of words, to change the function of the utterance (declarative vs. interrogative), to signify attitude and emotion, etc. These changes in fundamental frequency are related to the rate of vibration of the vocal folds. An increase in subglottal pressure and an increase in the tension of the laryngeal muscles (thereby increasing the stiffness of the folds and reducing their local mass) result in an increase in the fundamental frequency. Prosody can be used grammatically, by producing a question with a rising pitch, and a statement with a falling/flat pitch. Intonation also helps to group words, especially in syntactically ambiguous phrases (e.g. a light house keeper). However, there is quite a complex interaction between stress, intonation, and duration in the case of syntactic ambiguities. At some word boundaries, such as between 'gray tie' vs. 'great eye', duration is the primary cue, although the ambiguity can also be resolved through differentiation in intonation.

Stress

Stress usually refers to relative prominence and increased intensity due to increased physical effort. Note, however, that stress is not perceived as loudness, and that intensity is not the only acoustic correlate of greater effort. An increase in stress often co-occurs with an increase in fundamental frequency. After all, both stress and intonation are brought about by an increase in respiratory effort and subglottal pressure. The minimum size of the unit of stress placement is the syllable. At the word level, stress on the first or second syllable is used to differentiate meaning (cf: 'DIgest' and 'to diGEST' or RECord and to reCORD). This is termed lexial stress. At the sentence level, stress is used to draw attention to certain words (emphatic stress) or to differentiate between two words (contrastive stress). Languages are spoken with a certain rhythm. English is a stressed-timed language due to the fact that stressed syllables occur at regular intervals. French and Japanese are referred to as syllable-timed languages because the duration of their syllables remains relatively constant. Stressed syllables have a higher intensity and fundamental frequency than their unstressed counterparts. Apart from these voice source factors, the timing of the articulatory movements of the vowels and consonants produces differences in duration, depending on the position of the utterance in a word or phrase. In many languages, stressed syllables are longer than unstressed ones. The acoustical correlates of stress (especially duration) are, however, language dependent: in Czech, there is hardly an increase in duration on stressed syllables.

Quantity

Quantity is the linguistic term for 'constrastive duration'. Each individual speech sound has an 'intrinsic duration' to specify its identity, but the duration of speech segments or larger units can yield contrastive information, depending on context or the phonological system of the language. For example, the duration of vowels in Finnish can be constrastively short and long. In Norwegian, a long vowel is followed by a short consonant and vice versa. In English, the duration of a vowel preceding a voiced consonant is 1.5 times greater than that of the same vowel preceding a voiceless consonant (sight=197 ms, side=297 ms; Peterson and Lehiste 1960). Vowel duration, and consonant duration to a lesser extent, is also affected by the number of syllables in a word, by syllable stress (where it interacts with fundamental frequency), and by the type of word (function words such as 'in' or 'on' are relatively short). Very often, the last syllable of a word or the last word of an utterance is longer than the same syllable or word in other positions. This temporal effect, called prepausal lengthening, results from a slowing down in the speaking rate and it signals phrase and sentence boundaries. At word level, duration can signal the syntactic structure of a sentence, e.g. 'a gray tie' vs. 'a great eye'. At a sentence level, relative duration does not affect the meaning of individual words.

However, significant changes in tempo may convey information about the mood of the speaker or the circumstances under which the utterance is made.

References

Cooper, W., S. Eady, and P. Mueller. 1985. Acoustic aspects of contrastive stress in question–answer context. Journal of the Acoustical Society of America 77. 2124–56.

Cutler, A., and D.R. Ladd. 1983. Prosody: models and measurement. New York: Springer-Verlag.

Cutler, A. 1991. Linguistic rhythm and speech segmentation. Music, language, speech, and brain, ed. by J. Sundberg, L. Nord, and R. Carlson. Houndmills and London: MacMillan.

Lehiste, I. 1970. Suprasegmentals. Cambridge, MA: MIT Press.

Lehiste, I. 1972. The timing of utterances and linguistic boundaries. Journal of the Acoustical Society of America 51. 2018–24.

Lehiste, I. 1996. Suprasegmental features of speech. Principles of experimental phonetics. Mosby-Year Book, 226–44. St Louis, MO: Mosby.

Maeda, S. 1976. A characterization of American English intonation. Cambridge, MA: MIT Press.

Peterson, G.E., and I. Lehiste. (1960). Duration of syllable nuclei in English. Journal of the Acoustical Society of America 32. 693–703.

Terken, J. 1991. Fundamental frequency and perceived prominence of accented syllables. Journal of the Acoustical Society of America 89. 1768–76.

ASTRID VAN WIERINGEN

See also **Soundwave Analysis**

Psycholinguistics

Psycholinguistics is the discipline of psychology that studies the mental mechanisms of language processing—speaking, listening, reading, and writing—in both a native and a second tongue. Psycholinguistics also studies the processes underlying the acquisition of language, how language processes break down in language pathologies such as dyslexia and aphasia, and how these processes relate to brain function.

Psycholinguistics borrows many of its theoretical constructs from linguistics. Levels of processing, distinguished in theories of language comprehension or language production, correspond to linguistic levels, such as semantics (meaning), the lexicon (vocabulary), syntax (sentence structure), morphology (units that make up words), phonology, and phonetics (sound systems). Furthermore, the processing units that these theories assume correspond to linguistic units, such as the phoneme, the syllable, the morpheme, and the clause. Psycholinguistic experiments provide information about the psychological reality of linguistic units and the way linguistic information is represented and processed in the mind of the language user.

A crucial difference between linguistics and psycholinguistics is the latter's focus on the mental *process*. For instance, in natural speech we can easily produce two words per second. How do speakers find these words so quickly in their (vast) mental lexicon, which contains at least 10,000 words? What happens to these processes when we are unable to say a word, but it is on the tip of our tongue? As another example, when listening to sentences, we can sometimes be led up the 'garden path'. A classic example is 'the horse raced past the barn fell'. Here, one builds up a structural representation of the sentence, but upon reading the last word ('fell'), it turns out to be wrong. How do

readers build up a structural representation? What went wrong with this process in the example?

The scientific inquiry into language processing began in the second half of the nineteenth century. Rudolf Meringer (1859–1931) started to collect speech errors. In contrast to his contemporary Sigmund Freud (1856–1939), he studied the linguistic, rather than psychological, properties of such incidents. Wilhelm Wundt (1832–1920) proposed the first theory of language production, and Gustav Aschaffenburg (1866–1944), by using the technique of word association, investigated the representation of word meaning. The Dutch ophthalmologist Fransiscus Donders (1818–1889) introduced the method of mental chronometry (measuring the time mental processes take). Reaction time studies are still the most common methodology in psycholinguistics.

Around the same time, the French neurologist Paul Broca (1824–1880) tested a patient who had lost his ability to speak. Monsieur Tan, as the patient was known, could produce only a single syllable ('tan'). After Mr. Tan's death, Broca examined the brain of this patient and discovered massive damage to an area in the front part of the brain, which is now known as Broca's area. Broca's discovery, and subsequent studies by the German neurologist Carl Wernicke (1848–1905), marked the beginning of the inquiry into the relationship between language and brain.

In the first half of the twentieth century, psychology was dominated by behaviorism. This approach rejected the notion of 'mental representation' and 'mental process' but focused on what is directly observable: behavior. An exponent of behaviorism was Burrhus F. Skinner (1904–1990), who published his *Verbal behavior* in 1957. In Skinner's approach, the language-acquiring child begins without any linguistic knowledge. It learns language by experience only. Sentences would be produced as associative chains, where one word would be the stimulus that triggered the next word.

A few years later, a young linguist from the Massachusetts Institute of Technology, Noam Chomsky (1928–), published an extremely sharp criticism of Skinner's book. Chomsky argued that Skinner's theory could not work—for instance, a child is exposed to far too few examples of sentences to learn the language by experience only. Instead, Chomsky argued that the child is born with a tacit knowledge of 'universal' grammar, and learning means tuning this grammar to its own language.

Around the same time, psychologists were abandoning behaviorism and became interested in linguistics again. In 1951, a conference was organized at which leading psychologists and linguists discussed the possibilities of 'merging' the fields. At this conference, the term 'psycholinguistics' was launched.

Research in the 1960s and 1970s embraced Chomsky's theory of transformational grammar. This theory provided representations of sentences (called deep structure and surface structure), and procedures that transformed these representations. However, little convincing evidence was found for a correspondence between linguistic transformations and psychological processes, and this approach was largely abandoned in the early 1980s. Research in the 1980s and the 1990s was characterized by three major developments. First, Jerry Fodor published his *The modularity of mind* (1983), in which he presented the theory that many cognitive processes are performed by dedicated processing systems (modules) that operate independent of other processing systems. A large number of studies in modern psycholinguistics have been and still are centered on this theory. Second, psycholinguists designed computer models of language processing. Seminal work in this area was done by James McClelland and David Rummelhart (1989), in particular, on the modeling of word perception. Third, technical advances made it possible to measure the activity of the brain while it was engaged in verbal tasks. Although still in its infancy, the use of such techniques can be considered very promising in studying the representation and processing of linguistic information in the brain.

What are the mental processes involved in using language? In the following paragraphs, a brief overview will be presented of theoretical issues in pivotal language processes: comprehending, speaking, and 'control' processes.

The first challenge in comprehending is to identify words. This is no easy task: the speech signal has to be distinguished from noise, which may distort it. The speech signal is also very variable. One way of dealing with this variation is 'categorical perception'. That is, tokens of a speech sound can be acoustically quite different, but they tend to be perceived as the same sound (e.g. as the speech sounds /b/ or /p/, but never something in between).

Another challenge for the listener is that in continuous speech, there are often no pauses between words. How then do listeners know where one word ends and the next one begins? Psycholinguistic research shows that this depends on the language: different languages provide different cues to word onset (e.g. the pattern of stress), and listeners use these cues.

Visual word recognition—reading words—poses its own problems. English is an alphabetic language (as opposed to, e.g. Chinese). However, the correspondence between printed letters or groups of letters and sounds is often irregular (e.g. in the word 'yacht'). This has led psycholinguists to distinguish two routes for reading: a lexical route, in which letters or groups of letters directly access a word, and a sublexical

route, in which letters are transformed into a phonological code that guides word recognition. The lexical route would be used for reading irregular words; the sublexical route would be used for reading 'nonwords' (e.g. 'climp'). Support for this theory comes not only from distinct patterns of reading disorder (dyslexia) but also from experiments in which the speed of reading irregular and regular words is measured.

Once words are identified, we construct a structural representation of the sentence, so that we can infer who did what to whom. This process is called parsing. Many studies of parsing have analyzed structural ambiguities. In a sentence such as 'the man hit the girl with the guitar', the guitar can be in the possession of the man (and used for hitting the girl), or it can be in the possession of the girl being hit. These interpretations correspond to different sentence structures. Studies with ambiguous sentences have revealed some of the processes involved in parsing. A very important debate centers on the question of what kind of information is available to parsing. In some views, only syntactic information is used. According to other views, parsing is influenced by a variety of factors (e.g. plausibility, or the frequency with which a word occurs in a certain structure).

Sentence understanding does not stop with parsing. We have to integrate the word meanings and the sentence structure into a sentence meaning, and we have to relate this meaning to prior discourse. In one study, brain responses were measured to sentences such as 'I spread my bread with socks', and it was discovered that such stimuli evoke distinct patterns in the electroencephalogram signal. This finding has proved useful for studying these integration processes.

Speaking starts with the decision to engage in a communicative act. This is based on the speaker's intentions, background knowledge, and a mental model of previous dialogue. On the basis of these sources of information, the first step in speaking is the generation of a message. This is the interface between thought and language: the message is not yet language, but it is specific to language. In most theories of language production, the message contains concepts for the things and actions the sentence is about, perspective (is 'in front of' relative to you or me?), and focus (which information is given, and which is new?).

The next step consists of formulation of this message: retrieval of the words and construction of a sentence. There is much evidence that word retrieval consists of two steps: a process that is based on meaning and a process that is based on form. This distinction is based on research that analyzed collections of speech errors and research that measured reaction times in word production. In these tasks, so-called distracter words with either a form or a meaning relation to the target word have different effects. A major issue

is whether these data also imply that there are two separate lexical representations.

Another task of formulation is the construction of a sentence structure. A key finding is the observation that sentence structure is 'persistent'. Speakers have a tendency to reuse a structure that has been recently used before, even if the two sentences do not share any meaning, words, or prosody. This finding strongly suggests that structured representations of sentences are psychologically real. It also provides a technique for determining how sentences are mentally represented.

A final task of formulation is to retrieve pronunciations. A major problem here is that words in real speech do not correspond in a one-to-one fashion to words in their citation form: in connected speech, sounds from one word are often 'merged' with those of another (e.g. 'give it' will often be spoken as 'gi-vit'). Another issue is whether certain representations that are distinguished in theoretical phonology are psychologically real.

The earlier discussion focused mainly on content-directed processes in language use. Yet, there is increasing interest in control processes, which determine when the content-directed processes should engage, and in mental resources, such as memory and attention, which are the 'fuel' on which the content-directed processes run.

First, language is predominantly used in interaction with others. There is a growing body of evidence that partners in a dialogue coordinate with each other: that is, they converge on using the same linguistic structures, such as words, syntax, and description schemes, to denote the same things in the external world. How and why we accomplish such coordination is an important item on the agenda for study.

Second, the language processes may contain errors. Fortunately, there are self-monitoring processes dedicated to intercepting errors and 'repairing' them. How these monitoring processes operate—and how it is possible that they do the job as fast as they do—is, at present, a matter of debate.

Third, the language user is constrained by factors such as attention and memory. In recent years, there has been an increasing interest in how limitations in such 'resources' affect processing. Debate here centers on the nature of such resources: whether they are comparable to the same short-term memory in which one briefly stores the digits of a phone number, or whether there are specialized resources for language processing.

References

Carrol, David. W. 1999. Psychology of language. Pacific Grove, CA: Brooks/Cole.

Harley, Trevor A. 1995, 2001. The psychology of language: from data to theory, 2nd edition. Hove, UK: Psychology Press.

Fodor, Jerry A. 1983. The modularity of mind. London, England and Cambridge, MA: MIT Press.

Gernsbacher, Morton-Ann (ed.) 1994. Handbook of psycholinguistics. San Diego, CA: Academic Press.

Levelt, Willem J.M. 1989. Speaking: from intention to articulation. London, England and Cambridge, MA: MIT Press.

Brown, Colin M., and Peter Hagoort (eds.) 1999. The neurocognition of language. Oxford, UK: Oxford University Press.

Dijkstra, Ton, and Koenraad De Smedt (eds.) 1996. Computational psycholinguistics. London, UK: Taylor & Francis.

McClelland, James, L, and David E. Rummelhart. 1989. Parallel distributed processing: explorations in the microstructure of cognition, Vol. 2: Psychological and biological models. London, England and Cambridge, MA: MIT Press.

ROBERT J. HARTSUIKER

Q

Quantification

A sentence such as *John is happy* consists of a subject (*John*) and a predicate (*is happy*). The predicate attributes the property it describes ('being happy') to the individual the subject refers to ('John'). When the subject noun phrase contains a determiner (*a*, *the*, *some*, *every*, etc.), this element plays a crucial role in the predication process. The contrast in meaning between the sentences *A student is happy* and *Every student is happy* is caused by the different nature of the determiner that expresses the number or quantity of individuals to which the relevant property applies. We say, then, that these sentences differ in quantificational force. The first sentence states that there is (at least) one student who is happy in the situation under consideration. The second sentence states that every individual under consideration is happy. This semantic contrast can be expressed as a contrast in logical form, as defended by the philosophers Gottlob Frege and Bertrand Russell. In first-order logic, the determiner *a/some* is treated as an existential quantifier (meaning 'there is at least one individual for which it is true that…'), and the determiner *every* as the universal quantifier (meaning 'it is true for all individuals that…'). Quantifiers are assumed to bind a variable (x) in a logical statement or proposition. Going back to the above sentences, the paraphrases of their respective logical forms would be the following: *There is at least one individual* x *such that* x *is a student and* x *is happy* and *For every individual* x *if* x *is a student then* x *is happy*.

The first-order logic analysis of quantification has been criticized for positing that the constituents of a noun phrase (i.e. the determiner and the noun) contribute independently to the meaning of a sentence: the determiner quantifies over the whole proposition of which the noun is a part. This runs counter to the common linguistic analysis of sentential expressions. Another limitation is that the first-order logic analysis is not designed to capture the content of the extensive variety of natural language determiners. For example, it has been shown that the English determiner *most* is not expressible in first-order logic. The sentence *Most students are happy* does not mean that most individuals under consideration are students and they are happy (first-order logical meaning). Rather, it means that most of those individuals who are students are also happy.

Generalized Quantifier Theory developed during the 1980s, following the initial contribution of the philosopher Richard Montague, as an attempt to improve the shortcomings of traditional logical analyses of quantification in natural language. The functional nature of the subject—predicate relation that we described at the beginning is reversed. A predicate still expresses a property, but it is not treated as a function that applies directly to the subject (argument). Rather, a subject is viewed as a higher-order expression that expresses a property of the predicate or, equivalently, a set of properties. In general, noun phrases express generalized quantifiers (sets of properties). We can represent the sentences in the previous paragraph compositionally as *some student* (*happy*) and *every student* (*happy*), where the generalized quantifiers associated with *some student* and *every student* are respectively

interpreted as the set of properties that some student or every student has, and what the sentence asserts is that happiness is one of those properties. The sentence will be true or false depending on whether this is the case or not.

The meaning of a noun phrase such as *some student* can also be determined in a compositional fashion. The noun *student* expresses or describes, more technically 'denotes', a set of individuals and the determiner *some* relates it with another set of individuals, namely, that denoted by *happy*. This yields the following logical representation: a/*some* (*student*) (*happy*). Given the particular meaning of the determiner *some*, this expression is true if the intersection of the sets denoted by *students* and *happy* is not empty, i.e. if there is at least one individual who is a student and who is happy. Other determiner meanings can be characterized in a similar fashion: *every* (*student*) (*happy*) is true if the set denoted by *student* is a subset of the set denoted by *happy*; *no* (*student*) (*happy*) is true if the intersection of the sets denoted by *students* and *happy* is empty; *more than three* (*students*) (*happy*) is true if the number of individuals in the intersection of the sets denoted by *students* and *happy* is greater than three; *most* (*students*) (*happy*) is true if the number of individuals in the intersection of those two sets is greater than the number of the individuals in the intersection of the sets denoted by *students* and *not happy*, i.e. the number of those students who are happy is greater than the number of those students who are not happy; etc. In sum, a natural language determiner is a function with two arguments. The first argument corresponds to the noun denotation and the second argument to the content of the sentence verb phrase.

Natural language determiners also satisfy a series of constraints that set them apart from their logical counterparts. For example, all natural language determiners are conservative or live on their first argument—motivating the intuition that the determiner has a close bond with the noun. The sentence *Some sailors left* is equivalent to the sentence *Some sailors are sailors who left*. If we substitute any other determiner for *some* in this sentence, the equivalence still holds. The effect of this constraint is to make natural language quantification inherently restricted to the first argument of the determiner. Thus, in checking whether *some* (*sailors*) (*left*) is true, we do not have to consider those individuals who are not sailors. In other words, in processing the sentence *Some sailors left*, we do not first check who left and then determine whether any of these individuals are sailors; rather, we look at relevant sailors and determine whether some of them left.

Determiners are also characterized by the type of inferences that they license, namely, set-to-subset inferences or set-to-superset inferences. For example, the sentence *No students smoked* entails the sentence *No students smoked cigars*. The property denoted by *smoked cigars* is a subset of the property denoted by *smoked*, so the determiner *no* licenses set-to-subset inferences for its second argument. The same is true for its first argument: the sentence *No students smoked* entails the sentence *No female students smoked*, given that the property denoted by *female students* is a subset of the property denoted by *students*. The determiner function *some* has the opposite pattern and licenses set-to-superset inferences for its two arguments. Thus, we predict that the sentence *Some female students smoke cigars* entails both *Some female students smoke* and *Some students smoke cigars*.

Quantifiers can also be classified according to whether they can occur in an existential construction, i.e. a sentence that asserts the existence of something (*There is…*). Consider the structure *There is/are Q student(s) in the garden*. The determiners *some*, *three*, *no*, *fewer than five*, and *many* can be substituted for *Q* and occur in this construction, whereas the occurrence of determiners such as *every*, *most,* and *all but three* would make the sentence ungrammatical. The determiners that can occur in an existential construction are intersective, i.e. they express a relation of intersection of their two arguments. On the other hand, those determiners that cannot occur in an existential construction are characterized by expressing a relation of inclusion, such as *every*, or proportionality, such as *most*. Definite determiners, whether simple (*the*) or complex (*the ten…*), demonstrative (*this*) determiners, and possessive determiners (*my*) do not occur in existential sentences either. These determiners are all inherently context dependent and thus presuppose or do not assert existence.

When more than one quantifier occur in a sentence, a form of semantic interaction called scope arises. Scopal relations are determined by the different order of quantifiers in the semantic representation of a clause. For example, the sentence *Every student read a book* is ambiguous because it contains two generalized quantifiers: *every student* and *a book*. Under one interpretation, every student read a different book. Under the second, there is a unique book such that every student read it. This ambiguity is a genuine scope ambiguity. In the first reading, the scope order of the quantifiers is the one that respects the linear order of the noun phrases; i.e. the subject noun phrase is more dominant or has scope over the object: *every*

student > a book. The universal generalized quantifier *every student* takes scope over the existential generalized quantifier expressed by *a book*. The second reading is an inverse scope reading in that the scopal order of the quantifiers differs from the surface linear order. In this reading, the existential quantifier takes scope over the universal quantifier: *a book > every student*.

So far, we have considered only nominal quantification, i.e. quantification corresponding to the meaning of noun phrases. However, other elements may also contribute to the quantificational force of a sentence. Adverbs such as *always*, *sometimes*, and *often* are not merely temporal adverbs. In the majority of contexts, they behave as adverbs of quantification. For example, the sentence *Peter always drinks coffee* does not mean that Peter drinks coffee at every moment, but rather that whenever Peter drinks, he drinks coffee. Adverbs of quantification can thus be taken to express quantification over situations.

Interrogative generalized quantifiers are *wh*-phrases such as *who*, *what*, *where*, and *which*. Interrogative determiners are also conservative. The interrogative sentence *Which students are rich?* is equivalent to *Which students are students who are rich?* Interrogative determiners are also uniformly intersective, because the meaning of an interrogative determiner depends on the intersection of its two arguments. For example, a complete answer to *which* (*students*) (*rich*) specifies, in a situation, the intersection of the set of students and the set of rich individuals in that situation. Interrogative determiners and existential determiners such as *some* have a majority of their properties in common. From this point of view, it is not surprising that a significant number of the world's languages use the same lexical expression for interrogatives and existential determiners.

Sometimes, the quantificational force of a sentence cannot be associated with any overt element in the clause. The sentence *Lions are fierce* is usually interpreted as stating that most lions are fierce or that lions are normally fierce. To characterize the semantic properties of this sentence, the existence of hidden elements of diverse quantificational force has to be postulated. In the example under consideration, there is a hidden element of generic force, *Gen*, so the logical representation of this sentence would be *Gen* (*lions*) (*fierce*).

Nominal and adverbial quantification interact in apparently unexpected ways. This interaction arises mostly when an indefinite occurs in the scope of an adverb of quantification in relative constructions, conditional sentences, etc. In the sentence *Every farmer who has a donkey beats it*, the indefinite *a* lacks its typical existential force. The sentence does not mean that every farmer beats one donkey or other (at least one) that he has. Rather, the correct interpretation is that for every pair consisting of a farmer and a donkey owned by that farmer, it is also the case that the farmer beats the donkey. The indefinite seems to have universal force here; so we may conclude that the universal quantifier headed by *every* behaves as a binary quantifier here and associates both with the restriction of the universal noun phrase and with the indefinite one. The same pattern can be observed in the sentence *Always, if a farmer owns a donkey, he beats it*, where the indefinite noun phrases *a farmer* and *a donkey* seem to inherit the universal force from the adverb *always*. Some theories have concluded that indefinites lack quantificational force and inherit their apparent force from other surrounding elements in the clause. In general, adverbs of quantification are taken to transmit their quantificational force to all the indefinites that appear in their scope. This phenomenon is known as unselective quantification. Quantification thus covers a wide range of interesting phenomena and has triggered many debates among linguists and philosophers.

References

Benthem, Johan van, and Alice ter Meulen (eds.) 1985. Generalized quantifiers in natural language. Dordrecht: Foris.

Gärdenfors, Peter (ed.) 1987. Generalized quantifiers: linguistic and logical approaches. Dordrecht: Reidel.

Gutiérrez-Rexach, Javier. 1998. Semántica lógica y cuantificación nominal. Introducción a la teoría de cuantificadores generalizados. New Orleans: University Press of the South.

——— (ed.) 2003. Semantics: critical concepts. London: Routledge.

Hawkins, John. 1978. Definiteness and indefiniteness. London: Croom Helm.

Heim, Irene. 1982. The semantics od definite and indefinite noun phrases. Ph.D. dissertation. University of Massachusetts, Amherst.

Kamp, Hans, and Uwe Reyle. 1993. From discourse to logic. Dordrecht: Kluwer.

Keenan, Edward, and Leonard Faltz. 1985. Boolean semantics of natural language. Dordrecht: Reidel.

Lappin, Shalom. 1996. Handbook of contemporary semantic theory. Oxford: Basil Blackwell.

May, Robert. 1987. Logical form. Cambridge, MA: MIT Press.

Reuland, Eric, and Alice ter Meulen (eds.) 1987. The representatio of (in) definiteness. Cambridge, MA: MIT Press.

Szabolcsi, Anna. 1997. Ways of scope taking. Dordrecht: Kluwer.

JAVIER GUTIÉRREZ-REXACH

Questions

The term 'questions' in everyday usage refers to utterances inviting a response and that, when spoken, may have specific intonation patterns. Questions can be analyzed for their grammatical forms and also for their functions in discourse. Discussion of questions generally involves both these aspects. The analysis of questions can be used to relate linguistic findings to patterns of social behavior.

In English, questions can be identified and categorized on the basis of syntactic and phonological forms. Four types of questions in English can be distinguished:

(1)   Wh-questions

These are forms beginning with an interrogative element (*who, which, what, why, when, how*, and so on): *Where does Antony work*? *What can Antony do*? The subject (*Antony*) and either an auxiliary (*does*) or modal verb (*can*) are inverted in English. Wh- questions generally require more elaborate answers than those elicited by yes/no questions.

(2)   Yes/no questions

This broad category refers to questions to which the hearer is expected to respond either with yes or no. The following forms can be identified:

a. Inversion of the subject and an auxiliary or modal verb
   Does the race start here?
   Can we finish this now?
b. Statements with a question intonation

In these questions, the speaker presents an utterance in the form of a statement, but marks it as question by a rising and falling tone.
You will be sure to bring it back soon?

(3)   Alternative forms

In these forms, the speaker presents two or more alternative answers within the question.
Will the new administrator work here or on the first floor?

(4)   Tag forms

These forms consist of a declarative utterance followed by a question.
The computer is connected to the internet, isn't it?
Yes/no questions are said to either be neutral or biased toward a positive or negative orientation. For example, the use of the nonassertive form 'anyone' in *Has anyone left*? can be argued to render this question neutral in that it leaves open whether the response will be affirmative or not. However, the use of the more assertive form 'someone' in *Has someone left*? is argued to indicate speaker expectation of a positive response, whereas the use of a negative form hasn't in *Hasn't anyone left*? is argued to indicate a bias toward a negative response.

Generally, questions are understood to be based on presuppositions, or underlying assumptions. For example, the question *When does your brother finish college*? is based on the presuppositions that (1) you have a brother and (2) your brother is at college.

Questions can be examined in relation to their pragmatic functions, namely, their functions to request specific types of linguistic responses. A discourse-based analysis of questions involves a consideration of contextual factors, such as the situation, who knows what, and how the discourse unfolds. Identification of the functions is the primary interest, but identification of the forms that typically realize these functions may follow from this.

From a pragmatic viewpoint, a question can be identified as a question if it fulfills certain conditions: the act of questioning must be genuine, and the speaker must believe in the presuppositions of the question and desire to know the response. For example, a speaker (sitting in the driver's seat in a car with some passengers at the back) produces the following utterance: *Are you ready*? Without a response, the speaker then sets forth on the journey. This indicates that the driver does not desire to know the response. Also, the absence of any response from the passengers indicates that the hearers do not interpret the driver's utterance as seeking a response. Thus, despite the question-like form, the utterance would not be termed a question.

In discourse-based analyses, questions are often referred to as 'elicits' or 'requests', and they are categorized in relation to the specific functions they require of the hearer. One such category is elicits of, or requests for, information. In the following example (data taken from Schegloff 1972:107), A requires information about the location of an address from B.

A *I don't know where the—uh—this address is.*
B *Well, where do—-which part of the town do you live in?*

Another category in a discourse-based analysis is that of elicits of, or requests for, confirmation. These function to invite the hearer to verify or disconfirm the speaker's idea or proposition, as the following example taken from Coulthard and Brazil (1981:84) illustrates.

A *So the meeting's on Friday*
B *Thanks*
A *No I'm asking you*

This example also illustrates that identification of the exact function may be ambiguous and can only be resolved as the discourse unfolds. A's first utterance asks B to confirm that the meeting is on Friday. However, the function of this utterance is unclear, and B understands it as information being provided by A. B thus thanks A for the information. In utterance 3, A then makes explicit the function of the first utterance as an elicit of confirmation.

Other discourse functions of questions have been identified. Elicits of commitment require the hearer to promise further action (e.g. *You will come on Wednesday*?). Elicits of agreement require the hearer to verify that the idea proffered by the speaker is self-evidently true. The following example taken from Tsui (1992:107) illustrates:

(On a sunny day)

A *Lovely day, isn't it?*
B *Yes, beautiful!*

The analysis of questions according to function has been applied to the study of language use in specific situations. For example, analysis of teacher talk has been shown to reveal various didactic functions. These didactic functions include higher cognitive level questions requiring the hearer to manipulate bits of previously learned knowledge to create a response, lower cognitive level questions requiring the hearer simply to recall or recognize factual information, display questions functioning to check or test the student on information known to the teacher, and referential questions seeking information unknown to the teachers themselves.

Conversation analysts have analyzed questioning to identify rules underlying interaction. Questions are identified as the first part of a two-part sequence: question (Q) and answer (A). The person who asks a question has the option of speaking again to ask another question. Thus, there is a chaining rule providing for the sequence Q-A-Q-A-Q-A…. This approach to the analysis of questions has been used to relate linguistic findings to power structures in society. For example, the analysis of interaction may reveal one speaker asking more questions, thus directing the discourse and indicating asymmetrical power relationships.

References

Aijmer, K. 1996. Conversational routines in English: convention and creativity. New York: Longman.

Coulthard, R.M., and D.C. Brazil. 1981. Exchange structure. Studies in discourse analysis, ed. by R.M. Coulthard, and M.M. Montgomery, 82–106. London: Routledge and Kegan Paul.

Dillon, J.Y. 1990. The practice of questioning. London: Routledge.

Greenbaum, S., G. Leech, and J. Svartik. 1972. A grammar of contemporary English. London: Longman.

Schegloff, E.A. 1972. Notes on conversational practice: formulating place. Studies in social interaction, ed. by D. Sudnow, 75–119. New York: Free Press.

Tsui, A.B.M. 1992. A functional description of questions. Advances in spoken discourse, ed. by M. Coulthard, London: Routledge.

——— 1994. English conversation. Oxford: Oxford University Press.

van Lier, L. 1988. The classroom and the language learner. Harlow Essex: Longman.

HELEN BASTURKMEN

Quine, Willard van Orman

Quine is one of the most important analytic philosophers of the twentieth century, a fact eloquently attested to by the fact that in 1987 a supplement to the *Oxford English dictionary* recognized his importance as an eponym by including the entry 'Quinean' with the following definition: 'Of, pertaining to, or characteristic of Willard van. Quine or his theories'. At the time of his death in December 2000, Quine was the Edgar Pierce Professor Emeritus at Harvard, where he had pursued an active career spanning over half a century. His early work was heavily influenced by Rudolf Carnap and other logical positivists. But, as an empiricist and (from the 1960s on) an advocate of a naturalistic epistemology, he distrusted all talk of meaning

equivalence, which was a mainstay of Carnap's reductionist program.

A prolific writer, Quine is widely perceived as a philosophical iconoclast. His early essay 'Two dogmas of empiricism' (published in the 1953 collection) had great impact on the philosophical community. In it, Quine challenged the distinction between 'analytic' and 'synthetic' statements, one of the most fundamental and well-entrenched distinctions of modern philosophy—originally introduced by Kant but already foreshadowed in the work of Locke and Leibniz. According to Kant, an analytic statement was one whose predicate was contained in the concept of its subject. In the case of a synthetic statement, there was no such relation of inclusion between the subject and the predicate. Post-Kantian interpretation had resulted in an extension of the distinction to cover cases other than the simple subject–predicate statements originally envisioned by Kant and, with it, the key criterion for analyticity had become the idea of a sentence being true by virtue of its meaning. The idea had been embraced by the logical positivists of the Vienna Circle who had also extended its scope to include, among other things, all propositions of mathematics.

Quine's attack on the 'analytic/synthetic' distinction was based on his rejection of the notion of 'synonymy', which he thought was a philosophical mare's nest. Synonymy was suspect because it rested on the notion of meaning that too was, Quine argued, incapable of surviving critical scrutiny, notwithstanding the widespread use of the term in philosophical circles. Whatever heuristic value there was to the distinction was due to the fact that understanding a sentence was largely a matter of knowing what experiences would make a given sentence true, a claim long made by empiricists—the so-called analytic statements simply being, under this interpretation, those that have been confirmed on all known occasions.

Unlike the early Wittgenstein and the positivists, Quine saw philosophy as a natural ally of science. In his view, a scientific theory was an intricate web of ideas whose borders touched on concrete experience. No part of the web was immune to revisions in the light of new experience. Furthermore, revisions are never strictly localized but rather affect several parts of the web at once. This meant that individual sentences, considered in isolation, have no meaning, strictly speaking. In addition, Quine held that in talking about the world one characteristically moves from talking *in* certain terms to talking *about* them, that is, from so-called material mode to formal mode—a process that he famously referred to as 'semantic ascent'. On Quine's holistic view of meaning and

evidence, an entire network of beliefs and theoretical predilections is involved every time a scientific hypothesis is put to test. This means, he went on to argue, that no scientific hypothesis can be tested in isolation (this claim is often referred to as 'Quine–Duhem thesis'), just as no single sentence can be said to lend itself to an analysis of its putative meaning.

Skepticism concerning meaning and with it a whole array of assorted concepts such as property, proposition, and necessity as well as the very concept of a concept led Quine to unveil one of his most influential and controversial doctrines, namely the thesis of the 'indeterminacy of translation'. Quine argued that there can be no such thing as radical translation, i.e. translating, as it were, from scratch between two languages totally foreign to each other. A radical translator or interpreter will only have access to certain external facts, but these will be of no help when it comes to deciding between alternative candidates for the translation of an expression from one language to a radically different one. The impossibility of radical translation stems from what Quine called the 'inscrutability of reference' (later renamed 'indeterminacy of reference'). The native word 'gavagai' uttered by the informant as he points at a rabbit jumping past does not authorize one to hypothesize that the two words 'gavagai' and 'rabbit' are translation equivalents; a number of alternative hypotheses, such as that the informant might have been referring to specific rabbit parts or determinate rabbit movements, cannot be ruled out as equally plausible in light of the available empirical evidence.

At a broader level of application, Quine's position meant that, when different theories all seem equally adequate in explaining a given set of behavioral patterns, there can be no choosing among them on grounds of greater psychological reality or whatever. In an important paper titled 'Methodological reflections on current linguistic theory', Quine challenged Noam Chomsky's contention that the best theory is the one that is not only observationally and descriptively adequate, but explanatorily so as well. Quine insisted that, given a native speaker who presumably has a consummate command of his or her language and a hypothetical language learner whose linguistic performance is indistinguishable from that of the former, there is no sense in which the former may be claimed to *know* the language in any deeper sense than the latter. Quine's point was that, insofar as they are meant to be explanations for certain verifiable behavioral phenomena, the 'theory' that the language learner presumably brings to bear on the task of speaking the foreign language is just as good as the

one the native speaker is being claimed to have internalized. Accordingly, Quine dismissed Chomsky's celebrated notion of 'linguistic competence' as scientifically untenable. In the final analysis, Quine's objection to such notions had to do with his firm conviction in the irreducibly public nature of language. Toward the end of his career, he became more and more convinced that one can only learn language from experience with the world at large and that, therefore, there is no way linguistics can avoid embracing behaviorism.

Quine will be remembered as one of the most challenging and dynamic philosophers of the second half of the twentieth century. His hobbies included traveling and collecting stamps. He could never bring himself to use a computer, contenting himself instead with a 1927 Remington typewriter whose keyboard had to be modified to suit his special requirements of characters and type faces.

Biography

Quine was born in Akron, Ohio on June 25, 1908. He graduated from Oberlin College (Oberlin, Ohio) in 1930, and in 1932, barely two years later (an all-time Harvard record), received his doctorate from Harvard, with the famous English philosopher and mathematician Alfred North Whitehead as adviser. He traveled to Europe on a Sheldon Traveling Fellowship and spent the next year in Vienna, Prague, and Warsaw, studying, lecturing and meeting various members of the Vienna Circle, among them Philip Frank, Moritz Schlick, Alfred Tarski, A.J. Ayer, and Rudolf Carnap as well as Kurt Gödel. He started his teaching career at Harvard in 1936 as an Instructor in Philosophy and taught there ever since, except for four years in the US Navy during World War II, serving in Washington and working with cryptoanalysts trying to break the German submarine code. He was promoted to Associate Professor (1941), Professor (1948), and Edgar Pierce Professor (1956). He paid two visits to Oxford: in 1953–1954 as Eastman Visiting Professor, and in 1973–1974 as Savile Fellow of Merton College and Wolfson Lecturer. He was awarded 18 honorary degrees by international institutions, including University of Lille, Oxford University, Cambridge University, Uppsala University, University of Bern, and Harvard University. He received innumerable honorary fellowships and awards, including: Society of Fellows, Harvard University (Junior Fellow, 1933–1936; Senior Fellow, 1949–1978), American Academy of Arts and Sciences (fellow 1949 –), Harvard University (Chairman, Philosophy, 1952–1953), Association for Symbolic Logic (President, 1953–1955), Institute for Advanced Studies (Princeton, NJ, 1956–1957), American Philosophical Association (President 1957), American Philosophical Society, member (1957 –), Centre for Advanced Studies in the Behavioral Sciences (Palo Alto, CA, 1958–1959), British Academy corresponding fellow (1959–), Instituto Brasileiro de Filosofia corresponding member (1963–), Centre for Advanced Studies (Wesleyan University, Middletown, CT), Nicholas Murray Butler gold medal (1965), Columbia University (New York, 1970), National Academy of Sciences fellow (Washington, DC, 1977), Institut de France (1978), Norwegian Academy of Sciences (1979), F. Polacky gold medal (Prague, 1991), Charles University gold medal (Prague, 1993), Rolf Schock Prize (Sweden, 1993), and the Kyoto Prize (Japan, 1996). He retired as Edgar Pierce Professor Emeritus in 1978 from a teaching career in which his pupils had included not only influential philosophers like Donald Davidson but also the satirical songwriter Tom Lehrer and Theodore J. Kaczynski, the so-called 'Unabomber'. He died in Boston on December 25, 2000, aged 92.

References

Barrett, R., and R. Gibson, (eds.) 1993. Perspectives on Quine (Philosophers and their critics). London: Blackwell.

Davidson, D., and J. Hintikka (eds.) 1969. Words and objection: essays on the work of W.V. Quine. Dordrecht, Holland: D. Reidel.

Gibson, R. 1982. The philosophy of W.V. Quine: an expository essay. Florida.

Guttenplan, S. (ed.) 1975. Mind and language. Oxford: Oxford University Press.

Hahn, L. and P. Schilpp, (eds.) 1986. The philosophy of W.V. Quine. New York: Open Court.

Orenstein, A. 1977. Willard van Orman Quine. New York: Twayne.

Quine, Willard van Orman. 1934. A system of logistic. Cambridge, MA: Harvard University Press.

———. 1940. Mathematical logic. New York: Norton.

———. 1950. Methods of logic. New York: Holt.

———. 1953. From a logical point of view. Cambridge, MA: Harvard University Press.

———. 1960. Word and object. Cambridge, MA: MIT Press.

———. 1963. Set theory and its logic. Cambridge, MA: Harvard University Press, Belknap Press.

———. 1970. Philosophy of logic, Foundations of Philosophy Series, Englewood Cliffs, NJ: Prentice-Hall.

———. 1972. Methodological reflections on current linguistic theory. Semantics of natural language, ed. by D. Davidson and G. Harman. Doredrecht, Holland: D. Reidel Publishing Co.

———. 1974. The roots of reference. The Paul Carus Lectures, Vol. 14. LaSalle, IL: Open Court.

———. 1976. The ways of paradox, and other essays. Cambridge, MA: Harvard University Press.

———. 1977. Ontological relativity. New York: Columbia University Press.

QUINE, WILLARD VAN ORMAN

———. 1981. Theories and things. Cambridge, MA. and London: Harvard University Press.
———. 1989. Quiddities: an intermittently philosophical dictionary. Cambridge, MA: Harvard University Press.
———. 1992. Pursuit of truth. Cambridge, MA: Harvard University Press.
———. 1995. From stimulus to science. Cambridge, MA: Harvard University Press.

Romanos, G. 1983. Quine and analytic philosophy. Harvard: MIT Press.
Shahan, R. and C. Swoyer, (eds.) 1979. Essays on the philosophy of W.V. Quine. Oklahoma: University of Oklahoma Press.

KANAVILLIL RAJAGOPALAN

See also **Philosophy of Language**

R

Reading

Over the past several centuries, reading has gone from being a relatively obscure intellectual endeavor limited to a small elite segment of the population to a relatively common one. You are, in fact, reading at this very moment. Despite the fact that reading has become so common, and is something you probably do every day, picking a useful definition of reading turns out to be surprisingly difficult.

What Is Reading?

Reading a newspaper article surely counts as reading. But does scanning the stock market results to see how a favorite stock fared this week amount to reading? How about 'reading' a map? Is working through a mathematical proof, or a computer program, reading? Is it reading when you decode and pronounce a nonword such as BOPE?

Definitions of reading run the gamut from 'decoding printed symbols' to 'thinking with a book in front of you'. Differences in opinion about what constitutes a useful definition of reading arise from different answers to two fundamental questions. First, for something to be defined as reading, must it be unique to reading or can it be general to oral language comprehension? Reading almost always is done for the purpose of comprehending meaning, but most comprehension appears to be the same regardless of whether the message was initially processed through the eyes or the ears. Should a book on reading be mostly about language comprehension or should it be mostly about what is unique to reading as opposed to listening? The second fundamental question is whether comprehension is necessary for something to be considered reading. Studies of decoding printed letter strings sometimes use nonwords rather than real words to eliminate the effects of individual differences in vocabulary knowledge. If you can read, you will have little trouble working through a list of nonwords such as HEZ, RAF, POTE, and HINKER, and the processing that is done overlaps with that done when reading real words. But without comprehension, is this reading?

A pragmatic answer to the question of what is reading is to define reading as extracting meaning from print, and to acknowledge that it can be useful to study parts of the reading process such as aspects of decoding that are best viewed when people are asked to pronounce nonwords. In addition, it is important to acknowledge that successful reading requires the coordinated execution of decoding processes that are unique to processing print and comprehension processes that may be common to listening as well as reading.

What Is Being Read?

Humans have used spoken language for at least 100,000 years and perhaps quite a bit longer. In contrast, the earliest artifacts of writing, and hence reading, are cave paintings in southern France and northern Spain, that are approximately 20,000 years old. The paintings depict animals and occasional humans, and it is not clear whether they tell a story or are just pictures that serve some religious or magical purposes. A major advance in writing occurred around 5,000 years

ago in Sumeria. When goods were sent to market, they were accompanied by clay balls. Sealed inside the clay balls were small pieces of clay that represented the goods that were being transported. When the shipment reached the market, the clay ball would be broken open and the shipment would be inspected to make sure nothing had been diverted. Eventually, it became apparent that marks could be pressed into wet clay to represent the goods, saving the trouble of forming the balls filled with small pieces. This, of course, marks the birth of a real writing system.

Today, printed languages vary in terms of what they represent, with three major categories of writing. The Chinese writing system is called a logography. Each character represents a morpheme or minimal unit of meaning. *Beijing* means 'north capital", and is represented in writing by two symbols, one for north and one for capital. The Kana system of Japanese is a syllabary. Each symbol represents a spoken syllable. Finally, English, modern Korean, and many other written languages are alphabets, in which a small set of characters are put together to make a much larger number of words. Dividing the world's languages into three categories is an oversimplification. Thus, some sounds as well as meaning are represented in Chinese writing, and English is best described as a morpho-phonemic representation system as English spellings represent both sounds (phonemes) and (units of) meaning (morphemes). Consequently, the printed word HEALTH has a spelling that makes apparent its relation to the morpheme 'heal' as opposed to a strictly phonetic spelling such as HELTH.

How Do Children Learn to Read?

The obvious answer is that children are taught to read by teachers and parents. But a closer investigation of what is involved suggests it is immodest to assert that teachers or parents actually teach children to read.

Reading involves the coordinated orchestration of just about every process ever studied by cognitive psychologists and linguists. At best, teachers put print in front of children, and point out associations between print and sound or meaning, but the real work of learning to read is done by cognitive and linguistic machinery that is not readily accessible to teachers or learners. It is clear that learning to read an alphabetic system of writing is easier for children who have awareness and access to the sound structure of their oral language (i.e. phonological awareness) that is represented in writing. Thus, the spoken words 'cat', 'rat', and 'sat', have different initial sounds and identical medial and final sounds. To a child who is aware of these similarities and differences, the English system of writing will seem sensible in that these relations in sounds are represented in spelling. Cat, rat, and sat have different initial letters and identical medial and final letters. To a child lacking such an awareness, an alphabetic writing system will seem to be arbitrary and complex.

References

Adams, Marilyn. 1994. Beginning to read: thinking and learning about print. Cambridge, MA: MIT Press.

Crowder, Robert, and Richard Wagner. 1991. The psychology of reading. New York: Oxford University Press.

Perfetti, Charles. 1985. Reading ability. New York: Oxford University Press.

Rayner, Keith, and Alexander Pollatsek. 1989. The psychology of reading. Englewood Cliffs, NJ: Prentice-Hall.

Wagner, Richard, and Catherine McBride-Chang. 1996. The development of reading-related phonological processes. Annals of Child Development 12. 177–206.

Wagner, Richard, and Joseph Torgesen. 1987. The nature of phonological processing and its causal relations with reading. Psychological Bulletin 101. 192–212.

RICHARD WAGNER

See also **Handwriting; Literacy; Reading Impairment; Writing Systems**

Reading Impairment

Reading impairment, reading disability, and dyslexia all refer to a level of reading that is below expectations. Expectations for reading can be based on normative data from age-matched peers, or based on an individual child's oral language ability or cognitive ability.

Nature of Reading Impairment

Dyslexia is perhaps the most commonly used term to describe reading impairment, and surely the most misunderstood. The common view of dyslexia is that of a

visual perceptual problem that results in seeing mirror images of words or letters. Thus, individuals with dyslexia are reported to read WAS as SAW, or to confuse the letters 'b' and 'd'. The origin of this view is easily established. Children with reading problems typically became obvious to teachers and parents at the second-grade level, and indeed they can be observed reading WAS for SAW and confusing the letters 'b' and 'd'. But it turns out that normal beginning readers also make similar errors. The fact that words in English are to be read from left to right as opposed to right to left is arbitrary and must be learned. In addition, the letters 'b' and 'd' are both visually confusable and similar in sound (i.e. both stop consonants). Careful analysis of reading errors has shown that second-grade readers with dyslexia make no more reversal errors than do younger normal readers who are at the same level of reading. Thus, second-grade teachers observed only poor readers making reversal errors, whereas kindergarten teachers would know that such errors are quite common.

Another mistaken idea about reading impairment is that it results from erratic or inefficient eye movements. Reading requires highly sophisticated and coordinated eye movements characterized by ballistic movements called saccades and pauses called fixations. During the movements or saccades, little information is available to the eyes beyond a blur. Nearly all information is acquired during fixations. As you read the words on this page, your perception is that your eyes are moving smoothly across the page. This is an example of a situation where perception does not reflect reality. If you observe normal readers reading text by having them read directly across from you while holding a book low enough so you can observe their eyes, you will indeed see that the eyes move across the page in a series of small, but observable jerky movements.

Perform the same experiment on individuals with reading impairment and it will be apparent that their eyes move much more erratically, even moving in the wrong direction at times. Observations like these resulted in the belief that faulty eye movements were the origin of reading impairments, and also resulted in interventions based on eye-movement training. It turns out that this view has it backwards. Indeed, the eye movements of individuals with reading impairments are more erratic than those of normal readers, but the erratic eye movements are the by-product, not the cause, of the impaired reading. The eye movements of individuals with reading impairments do not move across the page as smoothly as do those of normal readers because they are having trouble reading the words. This also explains their greater frequency of

backward eye movements or regressions. Conclusive evidence was provided by careful studies in which normal readers were given material that was as difficult for them to read as is grade-level reading material for individuals with reading impairment, and individuals with reading impairment were given very easy reading material that they could read as well as normal readers could read grade-level material. Under these conditions, the eye movements of normal readers deteriorated to match the previously reported erratic eye movements of individuals with reading impairment, and the eye movements of individuals with reading impairment now looked normal. Additional confirmation came from the results of eye-movement training studies. Although eye-movement training did result in gains performance on eye-movement tasks outside the context of reading, reading performance did not improve.

For the vast majority of individuals with reading impairment, the problem appears to be based in language as opposed to the visual system, and is commonly compounded by ineffective instruction. Compared to reading-level matched controls, most individuals with reading impairment perform poorly on measures of phonological awareness and phonological decoding, and have fewer words that can be decoded by sight. Phonological awareness refers to an individual's awareness and access to the sound structure of an oral language. Phonological decoding refers to decoding words by sounding them out, as when one is asked to decode nonwords such as TANE. The underlying language problem for individuals with reading impairment is likely to be a subtle and not well-understood problem in forming accurate phonological representations, which in turn leads to poor phonological awareness and phonological decoding. Once beginning readers fall behind, they are exposed to reading instruction designed for more advanced readers, which provides little assistance, until they are finally identified as having a reading problem and more appropriate instruction is provided.

Distribution of Reading Impairment

Two important facts about the distribution of reading impairment are counterintuitive. First, despite the fact that boys outnumber girls by roughly four to one in classes and clinics that serve individuals with reading impairments, epidemiological studies in which all children are tested reveal that reading impairment rates for boys and girls are roughly comparable. For every boy with a reading impairment, there is likely to be a girl with an equally severe reading impairment. The overrepresentation of boys in classes and clinics appears to

result from a referral bias that arises because boys are more likely to be disruptive than girls, and hence more likely to be referred for evaluation.

Second, reading impairment is not something that one has or does not have. It is not an all or none phenomenon. The distribution of reading performance is continuous, with no obvious breaks or bumps in the tails of the distribution. Where one draws the line is arbitrary.

References

Blackman, Benita (ed.) 1997. Foundations of reading acquisition and dyslexia. Mahwah, NJ: Lawrence Erlbaum Associates.

Crowder, Robert, and Richard Wagner. 1991. The psychology of reading. New York: Oxford University Press.

Foorman, Barbara (ed.) 2003. Preventing and remediating reading difficulties: bringing science to scale. New York: York Press.

Shankweiler, Donald, and Isabelle Liberman. 1989. Phonology and reading disability. Ann Arbor, MI: University of Michigan Press.

Spear-Swerling, Louise, and Robert Sternberg. 1996. Off track: when poor readers become learning disabled. Boulder, CO: Westview Press.

Stanovich, Keith, and Linda Siegel. 1994. Phenotypic performance profile of children with reading disabilities: a regression-based test of the phonological-core variable-difference model. Journal of Educational Psychology 86. 24–53.

Wagner, Richard, and Tamara Garon. 1999. Learning disabilities in perspective. Perspectives on learning disabilities, ed. by Robert, Sternberg and Louise Spear-Swerling. Boulder, CO: Westview Press.

RICHARD WAGNER

See also **Reading**

Reference

The term 'reference' is ambiguously used to designate (a) the relation between a referring expression and that to which it refers, (b) that to which a referring expression refers (also known as the 'referent'), and (c) the act of referring to some extralinguistic entity or state of affairs, etc. by using a referring expression. Sometimes the terms 'denotation' and 'designation' are used interchangeably with reference in its sense (a)—'denotatum' and 'designatum' being the corresponding variants for the word in its sense (b).

Sense (c) introduces, over and above the referring expression and the referent, a third element, namely the speaker, into the picture. This in turn introduces the possibility that a speaker may succeed in referring to an object, although the referring expression she or he uses is, strictly speaking, inaccurate as a description of it—as when someone says 'That horse is agitated', pointing at a certain restless quadruped that, on closer inspection, turns out to be a mule. A speaker may also successfully refer to an individual whose identity is otherwise unknown to everyone including himself or herself as when someone says 'Smith's murderer is insane', meaning by the referring expression 'Smith's murderer' something like 'whosoever murdered Smith' rather than, say, 'that man in the dock over there, about to be cross-examined'. These cases show that speaker reference and denotation may not always coincide, raising the thorny issue as to which of them should be regarded as basic.

It is also important to distinguish among different kinds of referring expressions. 'That man over there', 'Smith', 'The President of the United States' are examples of singular definite expressions whereas 'horse', 'gold', etc. exemplify general expressions. Scholarly discussion about reference has tended to concentrate on singular definite expressions. Broadly speaking, theories of reference may be grouped under two conflicting approaches: descriptive vs. direct reference approaches.

On the descriptive view, a referring expression does its job by describing features or properties of the referent. The descriptive view hinges crucially on a distinction between *Sinn* (sense) and *Bedeutung* (reference) proposed by the German logician and philosopher Gottlob Frege in order to explain the difference in cognitive value between identity statements of the form '$x = y$' and '$x = x$', where x and y stand for referring expressions. Frege argued that two referring expressions may have different senses but nonetheless have the same reference (i.e. referent), thus making the former equation more informative than the latter. He also insisted that a referring expression can only have a referent provided it has a sense, although having a sense was not a sure guarantee that the term will have a referent (as in the case of 'unicorn' or 'square circle').

Bertrand Russell, who went on to spell out the description theory more articulately, took a narrower view of reference, maintaining that only 'logically

proper names' do the referring and they do it infallibly by naming objects known through 'acquaintance' that included sense-data, certain universals, and possibly ourselves. As for descriptions, he argued—contra Frege—that they are 'incomplete symbols', meaningful only in the sentential contexts in which they occur. Ordinary proper names like 'Bucephalus' and 'Mount Everest' are for Russell, as indeed they were for Frege too, disguised descriptions. So a descriptive phrase such as 'The present King of France' in the sentence 'The present King of France is bald' does not qualify as a genuinely referring expression, because a proper analysis of its logical form will show that what the sentence really says is that France currently has a (an only) king and that he is bald. Thus paraphrased, the singular definite description no longer has a referring function and its presence is explicated in terms of the existential quantifier and the variable within its scope as in

$$(\exists x)\ (Kx\ \&\ (\ (\forall y)\ (Ky \rightarrow y = x)\ )\ \&\ Bx)$$

where 'K' and 'B' stand for the predicates 'The present King of France' and 'Bald', respectively.

Russell's 'theory of definite descriptions' came under attack from Peter F. Strawson, who took up a line of reasoning initiated by Frege and, departing from it in significant ways, insisted that anyone who utters the sentence in question in fact presupposes that there is presently a king of France, rather than asserting it, as Russell's analysis implied. For Strawson, whereas referring expressions indeed have their senses, it is the actual *uses* of those expressions that do the referring. What specific horse is being referred to by the expression 'that horse' can only be resolved by examining an actual use (utterance) of the expression as part of, say, the making of a statement (and not merely the issuance of a sentence), along with the attendant set of actual circumstances. In other words, Strawson foregrounded the pragmatics of reference and claimed speaker reference to be paradigmatic, in opposition to Russell who was primarily interested in sense (a) or the semantics of the term 'reference'.

In contrast to the descriptive approach, direct reference approach harks back to a view taken by J.S. Mill in the nineteenth century. Unlike Frege and Russell, who argued that proper names had senses and that these coincided with the sense of corresponding definite descriptions, Mill had contended that they had no 'connotations' but only their 'denotations' (Mill's terms for sense and reference, respectively). In other words, they simply referred, period, by functioning as a label attached, as it were, directly to the object. Just how this direct contact with the referent is secured is an open question. One suggested answer is that there is a causal link between the use of a term and an original 'baptismal' episode of attaching the name to its

bearer. Not all direct theorists, however, subscribe to the thesis of causality.

Among modern defenders of the direct reference approach are Keith Donnellan, Saul Kripke, and Hilary Putnam. Highlighting the pragmatic aspect of reference, Donnellan argued that there are two distinct uses of definite descriptions, namely 'attributive' and 'referential' uses. In its referential use, a referring expression such as 'Smith's murderer' can succeed in picking up the intended referent, namely, say, the man in the dock awaiting the sentence, even though it may eventually turn out that the man in question is innocent of the charges. In other words, in its referential use a referring expression can refer *in spite of* its sense, contrary to what the descriptivist view predicts. In Donnellan's view, the Russell–Strawson debate was on a nonissue, since the former was thinking exclusively of the attributive use of definite descriptions while the latter was interested only in the referential use, but wrongly believing that the accuracy of the description was a necessary condition for the success of a referential use thereof. An interesting counter-argument to Donnellan's thesis was offered by John Searle, who claimed that the two uses, far from being mutually exclusive, can be explicated in terms of the number of 'aspects' under which a speaker is in a position to refer to the referent in question.

Kripke held that proper names are 'rigid designators', meaning that they have the same referents in all possible worlds. So it is wrong to think that a proper name such as 'George W. Bush' has the same sense as 'The President of the United States of America' because, although the two may indeed refer to the same individual in the actual world, they may not do so in some other possible world (say, the one in which the Supreme Court decision had favored Al Gore instead).

In the 1970s, Hilary Putnam developed the thesis of 'semantic externalism' whose central tenet is that, in many cases such as words referring to natural kinds like 'gold' and 'water', the reference of an expression cannot be thought of as a function of descriptions associated with it in the mind of the user. Thus, in Putnam's view, 'water' refers to the chemical compound H_2O in all possible worlds, irrespective of whether or not one is aware of that scientific truism.

Direct reference theories run into difficulties when called upon to explain the behavior of referring expressions in what are known as intensional or referentially opaque contexts as in 'Juan believes that the Malvinas rightfully belong to Argentina', which has—on one interpretation—a different truth value than the sentence 'Juan believes that the Falklands rightfully belong to Argentina' although 'Falklands' and 'Malvinas' designate the same group of islands (a fact that, as it happens,

Juan is not aware of). The usual answer that whatever difference there may be between the two sentences is not of interest to semantics has not satisfied critics. Causal theorists are also hard put to it to explain changes in the reference of a term over the years (as in the case of 'Yugoslavia' whose referent has significantly changed after successive wars) as well as natural kind terms that are posited rather than effectively attested (as in the case of a missing link to account for the development of human beings from apes).

References

Donnellan, Keith. 1966. Reference and definite descriptions. Philosophical Review 75.

Frege, Gottlob. 1949. Über Sinn und Bedeutung (1892). Readings in philosophical analysis, English translation by Herbert Feigl, ed. by Herbert Feigl and Wilfrid Sellars. New York: Appleton-Century Crofts, Inc.

Haack, Susan. 1978. Philosophy of logics. Cambridge: Cambridge University Press.

Kripke, Saul. 1972. Naming and necessity. Semantics of natural language, ed. by Donald Davidson and Gilbert Harman. Dordrecht, Holland: D. Reidel Publishing Company.

Lyons, John. 1977. Semantics, Vol. 1. Cambridge: Cambridge University Press.

Russell, Betrand. 1905. On denoting. Mind 14.

Searle, John. 1979. Referential and attributive. Expression and meaning. Cambridge, UK: Cambridge University Press.

Strawson, Peter F. 1971. On referring. Logico-linguistic papers. London: Methuen & Co.

<div align="right">KANAVILLIL RAJAGOPALAN</div>

Register

Broadly defined, 'register' refers to the way people use language in different situations. The term is often used as a full or partial synonym for style, speech variety or variation, field of discourse, and text type. It is the subject of sociolinguistic, linguistic, and applied linguistic analysis.

Most often, register is used to mean style. In that sense, it refers to the 'stylistic variety', or 'stylistic variation' that occurs in a person's speech or writing in different social contexts. Typically, different registers vary according to their degree of formality. For example, expressions like 'sure', 'no problem', and 'let's have a blast' are associated with informal conversational styles or registers. Utterances like 'in spite of our deliberate efforts to condemn fraud…', on the other hand, are used in formal written texts. In her *Grammar book*, D. Larsen-Freeman offers numerous examples of appropriate and inappropriate use of both formal and informal registers. In the case of phrasal verbs, 'put off', 'call off', and 'show up' are common in informal registers, whereas their one-word synonyms 'postpone', 'cancel', and 'arrive', reflect higher levels of formality. The same applies to relative adverbs. For example, '1950 is the year that I was born in' would occur in more formal texts than '1950 is when I was born'. A connector, such as 'moreover', normally occurs in formal written texts. It would be inappropriate in an informal conversational setting like this, 'Let's go to the beach. *Moreover, let's rent a boat'. It appears from these and other examples that there is an abundance of structures that reflect the subtle nuances of formality, both within a separate linguistic category and across different categories.

Linking register solely to levels of formality in speech and writing, however, would be an oversimplification. Along with levels of formality, scholars also study changes in register based on the speaker's awareness of a broader network of differences in situation, topic, addressee(s), or location. John Lyons emphasizes the complex sociolinguistic nature of register. He discusses it in the framework of context, style, and culture, and more specifically as a feature of 'communicative competence'. He relates it to the appropriate choice of language with regard to 'domain', i.e. a cluster of social situations constrained by a common set of behavioral rules. Within domain, Crystal and Davy identify 'province', i.e. occupational or professional activity with no reference to the individuals engaged in it. For example, the language used in advertising, science, law, and sports can be identified as being typical of the corresponding province, hence the use of 'love' in tennis and 'sentence' in law.

Fishman relates domain to 'subject-matter' on the one hand, and to 'locale' and 'role-relations' on the other. Examples of these would be 'family—home' and 'religion—church'. Within these subject-locales, he identifies certain role-relations, such as 'mother-to-son' and 'priest-to-parishioner'. Thus, a mother talks to her son at a different level of formality if the conversation is held at home or not. Similarly, her choice of language will be different if the topic was 'her son's drunk driving' and 'her son's recent award'. She would also choose a different register if she were to discuss these subjects with other people, at other places. Similar to Crystal and Davy's use of 'province', Quirk defines

register in relation to the 'field of discourse', or the activity in which a speaker is engaged. He categorizes speech varieties as dependent and interdependent according to 'field of discourse'. Speakers who retain regional features in their use of Standard English are examples of the former. Interdependent speech varieties, on the other hand, in a slightly broader use of Fishman's 'locale', are exemplified by topics that are associated with specific regions. Thus, discussions of baseball are in American English. Similarly, coaching is handled by speaking rather than by writing.

Linguistically, register could also be examined on the phonetic, morphological, syntactic, and semantic levels. Yet, it can be most fully analyzed on the 'text' level, e.g. it is suprasentential or above and beyond the sentence. In his *An introduction to functional grammar*, M.A.K. Halliday discusses register in relation to 'texture', or 'text structure', which comprises theme and focus, lexical cohesion, reference, ellipsis and substitution, and conjunction, all of which have a semantic, not grammatical structure. He points out some of the differences in 'theme and rheme' in different text types, such as narratives and instructions. In the following example of a narrative, the participant ('he') remains the topical theme: 'He went to many colleges but he didn't complete any. He always fell into some difficulty. He never hit upon the right thing'. In instructions that have stepwide structure, the theme, or given subject of one clause, becomes the rheme, or new object of the clause that follows it, e.g. 'Turn the stove on. Put a pot on the stove. Pour water into it. Add all ingredients to the water'. Thus, it would be inappropriate to use this kind of stepwide-structured register when writing an expository text, e.g. 'I have twin sisters. My parents love them. They buy them many toys. One such toy is a huge stuffed cat. It now occupies the couch'.

Halliday also defines register by different types of 'lexical cohesion and reference'. He notes that they are built through interlocking referential chain complexes that vary from one register to another and produce a certain global effect, e.g. 'A boy called John. . . John . . . he . . . the lad . . . him . . .' The global effect created by such overlapping referential chains is the source of the dynamic flow of discourse in narratives. According

to him, ellipsis and substitution, which are mostly found in dialogue, have a more local effect because of their shorter lexical reach. Typically, their reach extends over no more than a few consecutive moves, 'A: Would you like to come here? B: Could I bring a friend too? A: Sure. Which one? B: I'll know tomorrow. A: Is it still raining outside?' Such texts are related not so much by ideational as by interpersonal meaning. Halliday also defines the features of oral and written texts or registers. Generally characterized by lexical simplicity, oral texts become complex by being grammatically intricate. They use clauses to express relationships and to achieve dynamic complexity. On the other hand, written and technical texts, in particular, are lexically dense and statically complex because they use nominalization, or a lot of noun phrases, to express relationships. While they have simple clausal patterns, they pack ideas in nominal constructions, i.e. 'advances in technology' instead of 'technology is getting better'. Thus, nominalization, which helps to construct technical terms and develop step-by-step arguments, is often considered the most prominent feature of texts or registers of expert knowledge, prestige, and power.

References

Cheshire, Jenny. 1992. Register and style. International encyclopedia of linguistics, ed. by W. Bright. New York: Oxford University Press.

Crystal, David, and D. Davy. 1968. Investigating English style. London: Longman.

Downes, W. 1994. Register in literature. The encyclopedia of language and linguistics, ed. by R.E. Asher and J.M.Y. Simpson. Oxford: Pergamon Press.

Fishman, J.A. 1972. The sociology of language. Rowley, MA: Newbury House.

Halliday, M.A.K. 1994. Functional grammar. London, Melbourne and Auckland: Edward Arnold.

Larsen-Freeman, Diane, and Marianne Celce-Murcia. 1999. The grammar book, 2nd edition. New York: Heinle & Heinle.

Lyons, John. 1977. Semantics. London, New York and Melbourne: Cambridge University Press.

Quirk, Randolph, et al. 1985. A comprehensive grammar of the English language. London and New York: Longman.

Richards, Jack, et al. 1992. Longman dictionary of language teaching and applied linguistics. Harlow, Essex: Longman.

LILIA SAVOVA

Reichenbach, Hans

The philosopher Hans Reichenbach was a central figure of logical empiricism in the 1920s and 1930s. The starting point of his research were the philosophical dimensions and implications of relativity theory (*Relativitätstheorie und Erkenntnis a priori, 1920; relativity theory and cognition a priori*), but also more general issues in the

philosophy of space and time *(Axiomatik der relativistischen Raum-Zeit-Lehre (1924; Axioms of relativistic space–time), Philosophie der Raum-Zeit-Lehre (1928; Philosophy of space– time))*. In 1930, he founded the journal *Erkenntnis* (*Cognition*, later called *Journal of Unified Science*), together with Rudolf Carnap from the Vienna Circle. Furthermore, he developed a three-dimensional logic for the interpretation of quantum mechanics *(Philosophic foundations of quantum mechanics, 1944)* and worked on probability.

But linguists usually do not appreciate Reichenbach as a philosopher. In linguistics, he became famous for not more than 12 pages of his complete works, namely for §51 'The tenses of verbs' of his book *Elements of symbolic logic*, 1947.

According to Reichenbach, the tenses of verbs display token reflexivity. The tenses determine time with reference to the time point of the act of speech, i.e. of the token uttered. The time point of the token is called the point of speech (S), and the time point at which the event being talked about took place is called the point of the event (E). Using these two points means assuming three tenses, namely $E < S$ (the simple past in English), $E = S$ (the simple present), and $E > S$ (the simple future). But obviously, there are more than these three tenses in English (and in other languages as well). This is the reason why Reichenbach introduced a third parameter, namely the point of reference (R). What this parameter does becomes most easily clear in narratives. Compare the following passage from W. Somerset Maugham's *Of human bondage*:

> But Philip ceased to think of her a moment after he had settled down in his carriage. He thought only of the future. He had written to Mrs. Otter, the *massière* to whom Hayward had given him an introduction, and had in his pocket an invitation to tea on the following day.

The events related in the simple past perfect (*had settled down..., had written to Mrs. Otter..., had given him an introduction...*) are viewed as being completed from the perspective of a specific point of reference, which is constituted by those events that are related in the simple past (*ceased to think of her..., thought only of the future..., had in his pocket an invitation...*). Thus, the point of reference is a kind of perspective point.

With some tenses, two of the three parameters E, R, and S coincide. The simple past is such a case, with E and R being simultaneous, and both E and R being before S. With other tenses, the three parameters denote distinct times, e.g. in the case of the simple past perfect: E and R are both before S, but E is also before R. The different usages of the simple past and the simple past perfect in the quotation above show this quite clearly.

As for the simple tenses, Reichenbach proposes the following tables, in which the direction of time is represented as the direction of the line from left to right:

Past Perfect	*Simple Past*	*Present Perfect*
I had seen John	I saw John	I have seen John
–X–X–X–>	–X–X–X–>	–X–X–X–>
E R S	R,E S	E S,R

Present	*Simple Future*	*Future Perfect*
I see John	I will see John	I shall have seen John
–X–X–X–>	–X–X–X–>	–X–X–X–>
S,R,E	S,R E	S E R

If there is a progressive form, Reichenbach assumes that the event E covers a certain stretch of time. He calls these extended tenses. Here are two examples:

Past Perfect, Extended	*Simple Past, Extended*
I had been seeing John	I was seeing John
–XXXX–X—X–>	–XXXX—X–>
E R S	R,E S

As for adverbial modification, Reichenbach states that in the presence of positional adverbs like *now* or *yesterday*, only modification of R is possible. This can be illustrated with sentences like *I had met him yesterday*, where *yesterday* is the point of reference, but the meeting (E) has occurred at some time before *yesterday*. Here, R is the carrier of the time position. Therefore, Reichenbach speaks of the positional use of the reference point here.

On the whole, there are 13 possibilities of ordering the three time points E, R, and S. But for English, there are only six tenses in the traditional grammars. The following table summarizes the possibilities, Reichenbach's terms for them, and the traditional terms:

Possibility	Reichenbach's Term	Traditional Term
E_R_S	Anterior past	Past perfect
E,R_S	Simple past	Simple past
R_E_S	Posterior past	—
R_S,E	Posterior past	—
R_S_E	Posterior past	—
E_S,R	Anterior present	Present perfect
S,R,E	Simple present	Present
S,R_E	Posterior present	Simple future
S_E_R	Anterior future	Future perfect
S,E_R	Anterior future	Future perfect
E_S_R	Anterior future	Future perfect
S_R,E	Simple future	Simple future
S_R_E	Posterior future	—

With Reichenbach's terms, the position of R relative to S is indicated by past, present, future, and the position of E relative to R is indicated by anterior, simple, posterior.

As for the posterior future, it is used for instance in *I shall be going to see him*. Sentences like *I did not expect that he would win the race* express the posterior past—like posterior future this is commonly not classified as a tense.

Reichenbach has been criticized for assuming the point of reference because this often coincides with the point of the event or with the point of speech. Nevertheless, his work on tense is considered classic. Within the modern theory of tense, there are many reconstructions and adaptations of Reichenbach, e.g. Klein (1994), Ehrich (1992), Musan (2000), Comrie (1985), Thieroff (1992), and Fabricius-Hansen (1986), to mention just a few.

Biography

Hans Reichenbach was born in Hamburg, Germany, on September 26, 1891. He studied civil engineering, mathematics, physics, and philosophy at Technische Hochschule Stuttgart and Universities of Berlin, Göttingen, and Munich; Ph.D. (1915, Erlangen) for mathematical work about probability; and Habilitation (1920, Technische Hochschule Stuttgart) for philosophical work about relativity theory. He was assistant Professor, Technische Hochschule Stuttgart, 1920–1926; and Professor for Philosophy of Nature and Physics, University of Berlin, 1926–1933; he lost his teaching appointment when the Nazis seized power and emigrated to Turkey; he was Professor for Philosophy, University of Istanbul, 1933-1938. He moved to the United States, and was Professor for Theory of Science, University of California at Los Angeles, 1938–1953. He died in Los Angeles on April 9, 1953.

References

Comrie, B. 1985. Tense. Cambridge: Cambridge University Press.
Ehrich, V. 1992. Hier und jetzt. Studien zur temporalen und lokalen Deixis im Deutschen. Tübingen: Niemeyer.
Fabricius-Hansen, C. 1986. Tempus fugit: Über die Interpretation temporaler Strukturen im Deutschen. Düsseldorf: Schwann.
Klein, W. 1994. Time in language. London and New York: Routledge.
Musan, R. 2000. The semantics of perfect constructions and temporal adverbials in German. Habilitationsschrift, University of Berlin.
Reichenbach, Hans. 1920. Relativitätstheorie und Erkenntnis a priori.
———. 1924. Axiomatik der relativistischen Raum-Zeit-Lehre. Braunschweig: Vieweg.
———. 1928. Philosophie der Raum-Zeit-Lehre. Braunschweig: Vieweg.
———. 1935. Wahrscheinlichkeitslehre. Leyden: A.W. Stthoff.
———. 1944. Philosophic foundations of quantum mechanics. Berkeley: University of California Press.
———. 1947. Elements of symbolic logic. New York: Macmillan.
———. 1951. The rise of scientific philosophy. Berkeley: University of California Press.
Thieroff, R. 1992. Das finite Verb im Deutschen. Modus—Tempus—Distanz. Tübingen: Narr.

MONIKA RATHERT

Relevance in Discourse

Since discourse analysis is interested in stretches of text longer than a sentence or an utterance, it has to address the question of coherence—how the different parts of texts hang together. One crucial aspect of coherence will be relevance, how the meanings conveyed by one sentence can be related to or interact with information already in the reader's/hearer's mind, especially information already given in the preceding text.

The first philosopher to emphasize the importance of relevance to discoursal pragmatics was Grice (1975 [1967], 1989). His cooperative principle proposes four maxims that govern conversation in particular, namely quantity (give as much information as is required, no more), quality (only say things that are true for which you have evidence), manner (be clear, orderly, concise, and avoid obscure language), and relation (be relevant).

Sperber and Wilson (1995) and Wilson and Sperber (1987) developed Gricean theory by subsuming all these maxims under relevance. For them, communicative utterances carry the presumption of optimal relevance. Utterances are relevant if the information they convey interacts with the hearer's existing assumptions. These interactions, called contextual effects, can be either the strengthening of an existing assumption, the weakening/cancelation of an existing assumption, or the production of a contextual implication. In the

case of contextual implication, the information given by an utterance interacts with an existing assumption by deductive logic, to create a new assumption. One of Sperber and Wilson's examples is as follows:

A: Would you drive a Rolls Royce?
B: I wouldn't drive any expensive car.

B does not obviously answer A's question. However, if A already knows that a Rolls Royce is an expensive car, this assumption interacts with B's reply to produce the contextual implication 'B would not drive a Rolls Royce'. So, B's answer is still relevant to A's question.

The assumption supplied by the hearer 'a Rolls Royce is an expensive car' is called an implicated premise. That different contextual implications arise from different implicated premises is particularly important to the intertextuality of discourse. Intertextuality, put simply, means the way in which one text impinges on others (Kristeva 1974:59–60). One aspect, viewed from a relevance perspective, is that the information derived from one text can provide implicated premises and effect the contextual implications during the processing of a later text. For example, take the following text:

> It is difficult to see how any organism could remain healthy if feedback messages to the decision-making center were systematically stifled for long periods.

It will be processed and interpreted quite differently, more metaphorically, if the preceding intertext is a) a report showing that many politicians are devoting less and less time to their constituency surgeries/feedback sessions with the electorate rather than b) a feature article on the advances in the effectiveness of the pain-killing sprays used by football-players/athletes.

The relevance of utterances can be computed as a fraction:

$$\frac{\text{contextual effects}}{\text{processing effort.}}$$

All other things being equal, the greater the contextual effects the greater the relevance, and the greater the processing effort the less the relevance. We can see that contextual effects correspond to Gricean maxims of quality and relation; and processing effort corresponds to Grice's maxim of manner (and perhaps quantity). According to this formula, B's reply might not be seen as optimally relevant. B could have reduced the processing effort by simply replying 'No'. However, B's actual answer is likely to produce far more contextual effects than the answer 'No'. For example, if A knows that a Cadillac and a Lexus are also expensive cars, then B's actual reply will generate the contextual implications 'B would not drive a Cadillac or a Lexus'.

Like Gricean pragmatics, relevance theory opposes a code-based theory of communication. A successfully decoded message may be intelligible, as B's answer

above is to A, but to function as relevant in discourse it has to be interpretable, which may involve implication or inferencing. So Grice and Sperber/Wilson prefer an inferential account of communication. In this, coding and decoding using conventional signs are never sufficient for communication, and sometimes even unnecessary. If you ask whether I enjoyed my skiing trip, and I raise my broken leg covered in plaster, this might communicate the answer 'no'. But such an action has no conventional coded meaning. In inferential theories, utterances give evidence for a hypothesis about an intended meaning, often an implied one, as in the examples above.

However, relevance theory goes further than Grice, in maintaining that pragmatic inferential processes are important in making explicit the full propositional content of the utterance, even before implications are made. For example, imagine John has cooked dinner, Mary is sitting in the dining room reading the newspaper, and John comes in and puts the two plates of food on the table. He then says

> It'll get cold.

In order to make the prepositional content complete, we have to infer what 'it' refers to (for example, Mary's dinner rather than the newspaper), disambiguate 'cold' (to mean 'low in temperature' rather than 'experiencing low temperatures'), and narrow down the time reference of 'll' (to mean 'will soon'). This process, an example of what is termed explicature, will give us the relatively complete proposition

> Mary's dinner will soon become low in temperature

This is now intelligible as a complete message. But in order for it to be interpretable, it would have to generate a relevant contextual implication. If it interacts with Mary's existing assumption

> John wants Mary to eat her dinner while it is still hot.

the resulting contextual implication would be

> John wants Mary to come and eat her dinner very soon

(cf. Sperber and Wilson 1995:176–93).

Relevance theory was seen by Wilson and Sperber as a complement to Chomskyan linguistic theory, which deals with semantics but fails to account for the pragmatic meanings generated in discourse. Chomsky is famously uninterested in sociolinguistics. But, as early critics of relevance theory pointed out (Clark 1987), without some theory of social purposes the notion of relevance is rather vacuous. Goatly (1994, 1997) argues that in order to develop as a tool for discourse analysis, relevance theory needs developed notions of social contexts such as the Hallidayan

concepts of register and genre. However, Blass (1990) has argued that relevance theory gives a more satisfactory account of cohesion and coherence than the model of Halliday and Hasan (*Cohesion in English* 1976).

Relevance theory has been interestingly applied to different discourse genres, including work by Tanaka (1994) and Forceville (1994) analyzing advertisements, Campbell (1992) and Moeschler (1989) on argumentation, Wilson (1990) on political language, and Mayher (1990) for language in education.

References

Blakemore, Diane. 1992. Understanding utterances: an introduction to pragmatics. Oxford: Blackwell.

Blass, Regina. 1990. Relevance relations in discourse: a study with special reference to Sissala. Cambridge: Cambridge University Press.

Campbell, J. 1992. An applied relevance theory of the making and understanding of rhetorical arguments. Language and Communication 12(2).

Clark, H. 1987. Relevance to what? Behavioural and Brain Sciences 10(4).

Forceville, Charles. 1994. Pictorial metaphor in advertising. London: Routledge.

Goatly, Andrew. 1994. Register and the redemption of relevance Theory: the case of metaphor. Pragmatics 4(2).

———. 1997. The language of metaphors. London: Routledge.

Grice, Paul. 1975 [1967]. Logic and conversation. Syntax and semantics, Vol. 3: speech acts, ed. by Peter Cole and Jerry Morgan. New York: Academic Press.

———. 1989. Studies in the way of words. Cambridge, MA: Harvard University Press.

Halliday, M.A.K., and R. Hasan. 1976. Cohesion in English. London: Longman.

Kristeva, Julia. 1974. La revolution de langage poétique. Paris: Seuil.

Mayher, J. 1990. Uncommon sense: theoretical practice in language education. London: Heinemann.

Moeschler, J. 1989. Pragmatic connectives, argumentative coherence and relevance. Argumentation 3.3. Dordrecht: Kluwer.

Sperber, Dan, and Deirdre Wilson. 1995. Relevance: communication and cognition, 2nd edition. Oxford: Blackwell.

Tanaka, Keiko. 1994. Advertising language: a pragmatic approach to advertisements in Britain and Japan. London: Routledge.

Wilson, Deirdre, and Dan Sperber. 1987. An outline of relevance theory. Notes on Linguistics 39.

Wilson, J. 1990. Politically speaking: the pragmatic analysis of political language. London: Blackwell.

ANDREW GOATLY

Rhetoric and Linguistics

No discussion regarding the interrelationships between these two disciplines can proceed without a thorough understanding of what their content encompasses, the key to the alliance. The roots of rhetoric date back to the work of Aristotle (384–322 BCE) and the fifth-century sophists (460–380 BCE) such as Protagoras and Gorgias, who in their dialogues about the relationship of truth and reality, and the nature and purpose of spoken discourse consistently demonstrated that, inevitably, for these scholars there was a 'philosophical and pedagogical concern with language' (Gillam 1998:15).

Like linguistics, rhetoric has exhibited organic changes in focus, orientation, and scope over time, evident in the meaning of the word rhetoric, which has acquired meanings that are 'irreducibly multivalent' (Fleming 1998:174). Often associated in modern lay thought with the political rhetoric of 'deceit', rhetoric in academia has moved from being a 'pejorative to an honorific term' (Fleming 1998:169), a return to the classical art of persuasion. Modern uses of the term have broadened the scope of rhetoric to include 'virtually all humanly created symbols from which audi-

ences derive meanings' (Foss et al. 1999:6), an influence triggered by the modern rhetorician Kenneth Burke, who envisioned rhetoric to be 'symbolic inducement' (Burke 1966:296) for the purposes of either cooperation or competition. The site of location has shifted from public spheres of debate such as law, politics, and religion to more private constructions of meaning, particularly the manner in which individuals make sense of their world, a shift that has resulted in the synonymous connotation that rhetoric has acquired in modern times with the term 'composition', the private creation of meaning via the most common symbol of persuasion, writing, in modern universities. In this entry, these terms are sometimes used synonymously. Although the focus in classical accounts is with the construction of truth, modern conceptualizations see rhetoric as providing an arena to question the privileging of some truths over others, in short, 'a tool of critique' to demonstrate that 'reality, belief and language are not lined up as unproblematically as people would like' (Fleming 1998:170), a shift prompting an additional return to antiquarian roots espousing the sophist tradition of Isocrates which stresses civic development

or *Arete* (virtue) on the part of the individual. In such a conceptualization, 'rhetoric is a study that in addition to imparting an art and guiding good practice, encourages critical and substantive reflection about the situated relations of discourse to reason, character, and community in human action' (Fleming 1998:184), a definition stressing the link with linguistics, particularly sociolinguistics.

The link between the two fields is often conceptualized as the applications of linguistic theory to rhetoric, an alliance that can be broadly divided into two periods: optimistic applications of linguistic theory to rhetoric from the 1950s to the mid-1980s— a period of 'a good deal of excitement' (Smitherman 1999:351)— to pessimistic speculations as to the actual applicability of linguistic theory in the early 1990s to the present. It is important to stress that even though the literature seems to insist that the influence was unidirectional, linguistic research has also been influenced by studies in rhetoric. This might be attributed to the parallel development of composition-rhetoric and linguistics in the 1950s and 1960s as 'both fields sought to reinvent themselves and stake intellectual claim to distinct identities among the established disciplines of the academy' (Smitherman 1999:351). Consequently, the antiquarian interest in audience-influenced discourse styles saw linguistic applications in the work of Martin Joos (1961), who outlined five major audience-influenced styles of discourse.

Since language has and continues to be an integral part of rhetoric, especially in approaches labeled epistemic rhetoric, whose primary focus is on the private sphere of semantic generation, particularly the extent to which language simultaneously embodies and generates knowledge, linguists have constantly influenced and been influenced by the field. For instance, Kenneth Pike (1970) developed his theory of tagmemics in response to the argument that since the basis of language was the embodiment of knowledge, such a process could only proceed via access to reoccurring, repeatable units, the kind that language already offers. The dual relationship between these disciplines emerges in the unique conceptualization of Rogerian rhetoric, a persuasive technique based on conciliatory rather than adversarial argumentation. Similar links can be made with the work of Kenneth Burke (1966) whose model of language as action, 'an equipment for living' and 'a strategy for coping' (Warnock 1998:11) had linguistic parallels in the performative model of J.L. Austin's speech act theory.

The influence in the 1950s through the 1970s came from theoretical linguistics, particularly in the early phases, structuralism and, later, generative grammar, as well as applied linguistics, particularly the work of sociolinguists. One of the most overt influences in rhetoric-composition was in the form of the 1974 National Council of Teachers of English (NCTE) statement on 'Student's rights to their own language', a document first drafted by the Conference on College Composition and Communication (CCCC). The 1970s split in linguistics between Cartesian/theoretical linguistics and applied linguistics prompted a shift in influence, with most of the influence coming from applied linguistics. Some of the most commonly cited linguists at this time were Charles C. Fries, Kenneth Pike, Paul Roberts, and Noam Chomsky.

These linguistic influences prompted a shift from prescriptive accounts of written discourse to more descriptive orientations to student writing, as well as an overt recognition that the cultural, linguistic, and rhetorical backgrounds of writers shaped their discursive practices. Other areas of interest in this period came from studies in contrastive rhetoric, whose propounder, Robert Kaplan (1966), was heavily influenced by the work in contrastive analysis in applied linguistics. The aim of such early as well as later theorists such as Ulla Connor was to demonstrate first language sociocognitive influences of speakers engaged in second language rhetorical practices, work resonating the cultural determinism inherent in the linguistic theories of Benjamin Lee Whorf. Recognizing the sociocultural basis of writing, linguists such as Muriel Saville-Troike and Donna M. Johnson have synthesized ethnography of communication approaches with such conceptualizations and devised linguistically oriented rhetorical approaches whose basic premise is 'text as praxis' (Cai 1998:123). Applications of contrastive rhetoric in first language writing are manifest in the works of Mina Shaughnessy and David Bartholomae whose research stresses that student errors in writing may be reflections of systematic sociocultural patterns rather than random performance slips, an approach heavily adopted by theorists in basic writing paradigms such as that developed by Patricia Bizell and Bruce Herzberg who urge for the teaching of effective rhetorical strategies that permit both inter- and intracultural communication.

Research in the 1980s saw a deep alliance between the fields and 'scholarship influenced by Pragmatics, Discourse Analysis and Functionalist Linguistics appeared with some frequency in composition journals and monographs' (Kimball 2000:180). Shirley Brice Heath's (1983) *Ways with words*, an ethnography of speaking approach, influenced rhetoricians interested in issues of literacy and written composition then and even today. Other connections to linguistics came in the alliance with stylistics. This marriage of disciplines saw a culmination in the publication of two books, one by Raskin and Weiser (1987) of *Language and writing: applications of linguistics to rhetoric and*

composition, a book demonstrating pragmatic applications of linguistics in the teaching of rhetoric, and the other by Beale (1987) of *A pragmatic theory of rhetoric*, a rhetorical model based on J.L. Austin's Speech Act Theory (1962) since 'classical concepts of rhetoric square nicely with speech act theory, making them mutually illuminating' (Winterowd 1987:282). One can say that this was the last happy alliance.

By the early 1990s, institutional changes that separated English and Linguistics departments at a number of universities triggered pessimistic feelings on the part of rhetoricians concerning the utility of linguistic analysis, a sentiment often voiced in articulations that the scientific method of linguistics, particularly 'the atomistic nature of most linguistic analysis' (Roy 1991:580), was in contradiction to the humanistic method of rhetoric, a difficulty accentuated even further by the metalanguage adopted by linguistics that prompted Alice Roy (1991), in a classic review of four studies on the alliance, to describe the link as 'problematic' (580). Recent rhetoricians have commented that this pessimism is unwarranted, a consequence of a misapplication of linguistic theory, a view based on 'limited and inaccurate views of linguistics' (Kimball 2000:181). Utilizing research from Gary Olson and Lester Faigley, who review Chomskian applications to rhetoric concluding that 'It's very rare that you ever get a free ride from some other field' (34), prompts Kimball (2000) to call for a renewed interest in the applicability of linguistic theories to rhetoric today.

Obvious examples of such renewed interest come in discourse theories in rhetoric propounded by Kinneavy et al. (1990), as well as social constructivist research in the field by theorists such as Susan Jarratt and Lynn Worsham, Lester Faigley, and James Berlin, all of whom view language to be a complex, pluralistic signifying practice that simultaneously constructs realities as it presents it—a rhetorical paradigm rooted in Roland Barthes's (1968) theory of social semiotics, an approach emphasizing the social and ideological underpinnings of symbol use. These conceptualizations show heavy influences from linguistic research in critical discourse analysis, particularly social semiotics.

Renewed interest has occurred in the area of generative rhetoric, a paradigm originally devised by Christensen and Christensen (1976), who, influenced by generative linguistics, argued that texture in discourse proceeded via syntactic elaboration, particularly via modification and addition. There has been a recent pedagogical push for more overt sentence combining instruction in writing. An exponent, Kimball (2000) argues for this approach to be utilized in the revision stage of writing. Other current applications of linguistics are in genre theory, a theoretical paradigm strongly influenced by the works of M.A.K. Halliday's

systemic functional linguistics. Researchers such as Susan Jarratt and Lynn Worsham, Lester Faigley, and James Berlin push for a view of written rhetoric as situated social action, arguing that the only means to empower neophyte writers from being silenced is to educate them about the genres of powered discourse. Recent reapplications of early psycholinguistic and sociolinguistic models have manifested themselves in whole language teaching approaches to rhetoric, which emphasize linguistic socialization within and not apart from reading, writing, listening, and speaking. Other perspectives in genre theory have influenced rhetorical theories such as the literacy-orality theories of Walter Ong, an extension of Whorfian linguistics situating the source of writer difference in oral vs. analytical cultures, and/or individuals, an outcome of which may be a difference in cognitive complexity. Although this approach has been heavily critiqued in rhetorical and linguistic circles, applications of genre differences in spoken and written discourse have been recently cited by Kimball (2000) as an arena for research. Using the linguistic analyses of linguists such as M.A.K. Halliday, Wallace Chafe and Danielwicz, and Douglas Biber, she argues that some of the difficulties that novice rhetoricians face in this era may be an inability to distinguish between the features of this genre, a knowledge of which is presumed by the academy, and oral discourse. Returning to the unity of the disciplines in classical antiquity has caused some researchers in linguistics to recreate an alliance that benefits both pragmatics and rhetoric.

With the renewed interest in applied linguistic theories, the alliance between rhetoric and linguistics looks promising, a possibility accentuated by the fact that more and more students and faculty in composition-rhetoric and linguistics departments have some exposure to each other. What makes the synthesis even more viable is that the area of interest in both rhetoric and linguistics from antiquity to the present has been questions of language. In both fields therefore, we have to continue to see the product of language as a 'subject of study rather than an object of intervention' (Kimball 2000:188). With this said, one can only predict more of an alliance in this century.

References

Austin, J.L. 1962. How to do things with words. Oxford: Oxford University Press.

Barthes, Roland. 1968. Elements of semiotics, transl. by Anetto Lavers and Colin Smith. New York: Hill and Wang.

Beale, Walter H. 1987. A pragmatic theory of rhetoric. Carbondale: Southern Illinois University Press.

Burke, Kenneth. 1966. Rhetoric and poetics. Language as symbolic action: essays on life, literature and method, 295–307. Berkeley: University of California Press.

Cai, Guanjun. 1998. Contrastive rhetoric. Theorizing composi-tion: a critical sourcebook of theory and scholarship in con-temporary composition studies, ed. by Mary Lynch Kennedy, 52–6. Westport, CT: Greenwood Press.

Christensen, Francis, and Bonnijean Christensen. 1976. A new rhetoric. New York: Harper Press.

Fleming, David. 1998. Rhetoric as a course of study. College English 61(2). 169–91.

Foss, Karen, A., Sonja K. Foss, and Cindy L. Griffin. 1999. Feminist rhetorical theories. Thousand Oaks, CA: Sage Publications.

Gillam, Alice. 1998. Classical rhetoric. Theorizing composi-tion: a critical sourcebook of theory and scholarship in con-temporary composition studies, ed. by Mary Lynch Kennedy, 15–26. Westport, CT: Greenwood Press.

Heath, Shirley Brice. 1982. Protean shapes in literacy events: ever-shifting oral and literate traditions. Spoken and written language, ed. by Deborah Tannen, 91–117. Norwood, NJ: Ablex Publishing Company.

———. 1983. Ways with words: language, life, and work in communities and classrooms. New York: Cambridge University Press.

Joos, Martin. 1961. The five clocks. New York: Harcourt, Brace and World.

Kaplan, Robert B. 1966. Cultural thought patterns in intercul-tural education. Language Learning: A Journal of Applied Linguistics 17. 1–20.

Kimball, Sara. 2000. Linguistics and composition. Reforming college composition: writing the wrongs, ed. by Ray Wallace, Alan Jackson and Susan Lewis Wallace, 179–90. Westport, CT: Greenwood Press.

Kinneavy, James L., William J. McCleary, and Neil Nakadate. 1990. Writing in the liberal arts tradition, 2nd edition. New York: Harper.

Pike, Kenneth L. 1970. Tagmemics and matrix linguistics applied to selected African languages. Norman: Summer Institute of Linguistics of the University of Oklahoma.

Raskin, Victor, and Irwin Weiser. 1987. Language and writing: applications of linguistics to rhetoric and composition. Norwood, NJ: Ablex Publishing Corporation.

Roy, Alice. 1991. Four studies of linguistics and composition. College English 53(5). 580–6.

Smitherman, Geneva. 1999. CCCC's role in the struggle for language rights. College Composition and Communication 50(3). 349–75.

Warnock, Tilly. 1998. Burkean theories of rhetoric. Theorizing composition: a critical sourcebook of theory and scholarship in contemporary composition studies, ed. by Mary Lynch Kennedy, 9–13. Westport, CT: Greenwood Press.

Winterowd, Ross W. 1987. Literacy, linguistics and rhetoric. Teaching composition: twelve bibliographic essays, ed. by Gary Tate, 265–90. Fort Worth: Texas Christian University Press.

ANJALI PANDEY

Romanian

The Romanian language (*Limba română*) is a Romance language that has been continuously spoken in the east-ern part of the Roman Empire, including the Romanized provinces of Dacia, Southern Pannonia, Dardania, and Moesia, from the moment Latin was brought there until the present. The Romanian language has four main dialects: Daco-Romanian (or Romanian proper), Macedo-Romanian or Aromanian, Megleno-Romanian, and Istro-Romanian. The Aromanian dialect is spoken in Albania, Macedonia, Greece, and Bulgaria by scat-tered minorities, and the main area where Megleno-Romanian is spoken is in Greece. The Istro-Romanian dialect was spoken on the Istrian Peninsula of Croatia. The Daco-Romanian dialect ('Romanian') is spoken by more than 25 million people in Romania, Serbia, Croatia, Slovenia, Bosnia and Herzegovina, Bulgaria, Hungary, and Moldova, as well as in Romanian com-munities in the USA, Australia, Italy, Germany, etc. The variety of Romanian spoken in the Republic of Moldova (east of the Prut river) is officially called 'the Moldavian language'.

The earliest known written document in Romanian is a letter dating from 1521. The present spelling system was settled around 1860–1880. The basis of the standard Romanian language is a combination of the two main geographical varieties, the Moldavian (west of the Prut river) and the Wallachian (Muntenian), the latter playing the more important role in setting the standard. Given the language's relative isolation from the other Romance languages and its close contact with Hungarian and the Slavic languages surrounding Romania, the develop-ment of Romanian followed a different path from those of most other Romance languages.

One of the most important foreign influences came from the Slavic languages. As a liturgical language, Old Slav(on)ic (or Old Bulgarian) provided a relatively large number of words to Romanian. This situation, combined with the fact that until the nineteenth century the Cyrillic alphabet was used to write Romanian and with the geopolitical position of Romania, has misled many peo-ple into thinking that Romanian is a Slavic language and that Romania is a successor state of the Soviet Union. Another major influence in the development of the Romanian vocabulary came from Greek, especially dur-ing the Phanariot period (1711–1821). However, very few of the Greek borrowings are still in use today. Other foreign influences came from Turkish, Hungarian, and German. Turkish influence was prominent in Moldavia

and Wallachia during the Phanariot period, and Hungarian influence was prominent in Transylvania while it was part of the Austro-Hungarian Empire. The German elements used by the Romanian minorities of German origin also played a role in the diversification of the Romanian vocabulary. More recent foreign influences came from French in the first part of the twentieth century, while today massive English influence on the Romanian vocabulary can be easily observed.

Within the geographical area of Romania, there are three main varieties of the language: Moldavian, Wallachian (Muntenian), and Transylvanian. Additionally, in the Republic of Moldavia the inhabitants speak the Bessarabian variety of Romanian, which is characterized by the massive influence of Russian. All these varieties, however, are mutually intelligible. The standard pronunciation of Romanian is based on orthographic norms. In this respect, the principle generally promoted is that of identity between writing and pronunciation.

As far as the vocabulary is concerned, Romanian inherited c. 2,000 original Latin words, but Latin elements make up 60% of the Romanian vocabulary. Most of the original Latin words refer to parts of the human body and have several meanings, including metaphorical ones.

In contrast to the vocabulary, Romanian grammatical structure is almost entirely Latin. There are Latin prepositions (e.g. *cu* 'with') and conjunctions (e.g. *și* 'and'), Latin interrogative pronouns (e.g. *cine* 'who'), Latin endings for marking the plural, Latin verbal endings for marking person and tense, Latin pronouns, and Latin numerals from one to ten.

Romanian grammar involves ten parts of speech: six inflected (noun, article, adjective, pronoun, numeral, and verb) and four noninflected (adverb, preposition, conjunction, and interjection). The following grammatical categories all stem from Latin: gender, number, case, person, comparison, voice, mood, and tense. However, the development from Latin to Modern Standard Romanian generated several differences. Thus, for example, Latin had six cases, while Romanian preserves only five—nominative, genitive, dative, accusative, and vocative.

Gender in Romanian distinguishes masculine, feminine, and neuter (masculine and feminine genders are inherited from Latin, while neuter is a creation within Romanian).

The changes from Latin are typically simplifications (as in all Romance languages), which led to a reorganization of the vocabulary by reducing the complicated forms of the noun and adjective, by simplifying and rearranging the verbal system, and by simplifying the case system.

While the Romance languages usually have reduced nominal declensions, Romanian has three forms of noun inflection inherited from Latin. The declension of adjectives is identical to that of nouns, varying according to gender, number, and case. The pronominal declension is the richest and the closest to the complex Latin system of all the Romance languages. The Romanian verbal system is largely based on Latin. The most noticeable differences from the other Romance languages are the analytic future (*voi citi* 'I shall read') and the supine (*de citit* 'for reading'). Romanian is the only Romance language with a free sequence of tenses and a relatively free word order.

References

Constantinescu-Dobridor, Gheorghe. 1994. Sintaxa limbii române (The syntax of the Romanian language). Bucharest: Editura Stiintifica.

Coteanu, Ion. 1993. Gramatica de bază a limbii române (The fundamental grammar of the Romanian language). Bucharest: Garamond.

Daniliuc, Radu, and Laura Daniliuc. 2000. Descriptive Romanian grammar. An outline. Munich: Lincom Europa.

Doca, Gheorghe. 1996. Limba Română. Probleme de sinonimie și de cultivare a limbii (Romanian language. Synonymy and style issues). Bucharest: TUB.

Harris, Martin, and Nigel Vincent (eds.) 1990. The romance languages. London: Routledge.

Pop, Liana. 1993. Romanian with or without a teacher. Cluj: Echinox.

Posner, Rebecca. 1996. The romance languages. Cambridge: Cambridge University Press.

Serban, Domnica. 1986. Syntactic functions in universal grammar. A contrastive study of the subject in English and Romanian. Bucharest: TUB.

RADU DANILIUC

Rules vs. Constraints

The question of whether grammar should be modeled using rules or constraints presents a morass of terminological confusion in addition to some legitimate theoretical issues. In some work, e.g. Karttunen (1993), the terms '(declarative) rules' and 'constraints' are used interchangeably. There are even positions

such as Mohanan's (2000:146), who argues that the distinction is incoherent once we recognize that both rules and constraints express propositions. This article proposes a definition of 'rule' vs. 'constraint' that will surely be inconsistent with other usages, but which provides the basis for discussing competing models of grammar, focussing here on phonology.

The use of both rules and constraints requires a representational matching procedure (RMP) to determine whether a given rule or constraint P applies to a given representation q. Part of the statement of P must be a structural description, SDp. The RMP must evaluate q to determine whether it satisfies SDp. The RMP outputs two possible results: YES, q satisfies the structural description of P; or NO, q does not satisfy the structural description of P. In other words, the RMP defines the domain of the function P. In many theories, the RMP relies on subsumption—q is in the domain of P if SDp subsumes q—although there are other possibilities.

If the RMP determines that q is in the domain of P, we can now distinguish between a rule and a constraint. Consider a rule of the form $R(I) = O$, that is, R maps input I to output O. If q is in the domain of R (i.e. if the RMP determines that q matches I), then that part of q that is identical to I is rewritten as O yielding q': $R(q) = q'$. Thus, we define rules as functions from representations to representations, with both input and output built from the same representational primitives.

In a constraint-based system, if the RMP determines that q is in the domain of a constraint C, then an operation of constraint evaluation must determine whether q satisfies the constraint. If q satisfies C, constraint evaluation outputs the value NOVIOLATION; if q does not satisfy C, constraint evaluation outputs the value VIOLATION. These values then serve as input to another component of the grammar. In some models, if there is any constraint that q violates, then q is ungrammatical. In optimality theory (OT), in contrast, the evaluations of q for each constraint are fed to another component of the grammar, which compares q to competing representations. The output of the grammar typically does violate some constraints in OT, but they must be lower ranked than constraints violated by competing representations.

To reiterate: a rule is defined as a function from representations to representations; a constraint is defined as a function from representations to the set {VIOLATION, NOVIOLATION}. Note that a constraint-based system requires two evaluations of a given representation q—one corresponding to the RMP to determine whether a constraint C is relevant to q and one to the evaluation of whether q violates C.

Equipped with this working distinction between rule and constraint, we can now offer arguments to favor rule-based models of grammar, by pointing out a number of conceptual and empirical problems with constraints. Our first criticism of constraint-based models is related to issues of learnability, acquisition, and the nature of universal grammar (UG), i.e. the human language faculty. Many linguists, especially phonologists, have assumed that both UG and particular grammars contain constraints—either positive or negative conditions that, respectively, must or must not be satisfied by grammatical representations. However, such conditions *cannot* be learned by positive evidence. This is because a generative grammar generates an unbounded number of well-formed structures, but only a finite sample can be encountered by any point in acquisition. The learner may find a supposed ill-formed structure in the next sentence. For example, a putative constraint that all syllables must have onsets may be consistent with the data received up to a given point; however, there is no guarantee that the inductive conclusion consistent with the constraint will turn out to be correct. Therefore, such conditions could only be learned via negative evidence. However, it is generally accepted (see Marcus (1993) for convincing arguments) that negative evidence is neither supplied to the child with sufficient regularity, nor attended to by the child when supplied, to play a significant role in language learning. Thus, the constraints cannot be learned via positive evidence (for reasons of logic), nor through negative evidence (according to the empirical data from language acquisition). So they must be innate.

This conclusion follows from the premises, and it has led to implausibly rich versions of UG. In phonology, for example, innate constraints have been posited to account for the vast variety of phonological alternations seen in the languages of the world. The OT literature provides examples of constraints to account for voicing of initial obstruents in the second member of Japanese compounds, flapping in English, and so on, thus trivializing the notion of UG. Fortunately, the conclusion of innate constraints can be rejected: the problem lies with the assumption that UG, and also particular grammars, consist of constraints. By positing a set of representational and computational primitives, such as syllables and feet, precedence relations, identity evaluations, etc., and a learning mechanism that combines these primitives into rules that capture alternations, a learner can build a language-specific grammar based solely on positive evidence. By positing highly specific rules that conform to the data, the learning path is constrained without the use of constraints.

A cluster of arguments against constraint-based models relate to the fact that they typically depend on adopting a notion of (absolute or relative) ill-formedness

of linguistic representations. Ill-formed representations violate constraints, and this result is passed on to other modules of the grammar. An obvious objection to this approach is that there is a much simpler alternative to the idea that grammars generate ill-formed representations and then reject them as possible outputs—why not assume that the grammar just generates those representations that it outputs? The grammar itself does not need the notion of ill-formedness. Of course, a linguist may refer to a representation as ill-formed, in the sense of 'not generatable by a particular grammar *L*', but this description by a linguist should not be interpreted as a description of an actual property of *L*. Instead of accounting for the infinite set of representations that *L* does not generate, normal scientific practice seems to dictate that we account for what *L* does generate. Ill-formedness, or rather non-formedness, is thus a derivative notion.

Various theories of grammar, including OT and some versions of minimalism and its predecessors posit a mechanism that allows unconstrained generation of linguistic representations. In OT, this device is GEN which, given an input, generates the universal candidate set of possible outputs. In various syntactic theories, an analog to GEN is the 'free' concatenation of morphemes, or the 'free' application of operations such as *Move α*. A derivation which is thus generated will either satisfy certain constraints at PF and LF, the grammar's interface levels, and thus converge; or it will not satisfy those conditions and it will crash. Both the OT approach and the free-generation-with-interface-conditions approach in syntax are flawed in the following (related) ways.

First, it is easy to proclaim something like 'GEN generates any possible linguistic representation' or 'The syntactic component allows *Move α* to apply freely'. However, it is not clear what such statements mean. One could argue that the theory of grammar need not be computationally tractable, since grammar models knowledge and does not necessarily provide an algorithm for humans to produce speech output. However, there must be some mechanism that does generate particular linguistic representations, and it is not unreasonable to ask a theory of grammar to characterize the formal properties of this mechanism. It seems that any implementation of GEN or the syntactic component that incorporates *Move α* will have to be explicit about what it does. One way to achieve this is to be explicit about what the abstract grammar generates.

Second, the free generation-cum-filters model stinks somewhat of antimentalism. It basically says 'We don't care how the candidate forms are generated, as long as they are generated. One way is as good as the next, as long as they are extensionally (empirically) equivalent.' In defining I-language, a matter of 'individual psychology' as the domain of inquiry for linguistics, Chomsky (1986) argued convincingly that the fact that knowledge of language is instantiated in individual minds/brains means that there is necessarily a 'correct' characterization of a speaker's grammar.

Rejecting the view that the language faculty consists of constraints has methodological implications. Just as we do not want the language faculty to contain constraints against grammatical structure, models of the language faculty should not be described via constraints. If our current hypothesis concerning UG is stated only in positive terms, as statements of what grammars have access to or consist of, without prohibitions or constraints, we can achieve a more economical model. The positive terms are just those entities and operations (features, syllables, deletions, insertions (in phonology); and *Merge, Move* and feature-checking (in syntax)) which have been observed empirically or inferred in the course of model construction. When faced with a phenomenon that is not immediately amenable to modeling using existing elements of the vocabulary, scientific methodology (basically Occam's Razor) guides us. We must first try to reduce the new phenomenon to a description in terms of the vocabulary we already have. If this can be shown to be impossible, only then can we justify expanding the vocabulary.

Once one accepts that modules or processes of grammar must have a certain set of properties and that these properties ultimately must be describable in terms of a set of positive statements (a vocabulary) and can be incorporated into the structural descriptions of rules, it appears to be the case that a procedural, or rule-based approach to grammar that generates a sequence of representations constituting a derivation is to be preferred to a constraint-based, nonderivational theory. In other words, grammars can be understood as complex functions mapping inputs to outputs. A rule-based model just breaks the complex function into simpler components in order to understand the whole. A theory that incorporates GEN or *Move α* avoids the problem of characterizing the function that is the grammar.

A further objection to the notion of ill-formedness is the fact that the justification for evaluating representations as ill-formed is typically derived from some version of markedness theory. Depending on the formulation of a given constraint, either matching or failing to fulfill the condition specified by the constraint signals VIOLATION. For example, a constraint formulated as 'Don't have a coda' leads to an evaluation of ill-formedness for a syllable that *has* a coda. A constraint formulated as 'Have an onset' leads to an evaluation of ill-formedness for a syllable that *does not have* an onset. Relative and absolute ill-formedness or markedness

evaluations of linguistic representations are ascribed by linguists for grammar-external reasons, such as ease of articulation or perception, avoidance of ambiguity, etc., and thus have no place in a formal model of grammar—humans know about ambiguity, and can try to avoid it, but grammars do not know anything.

Marked or ill-formed structures typically are claimed to have at least one of the following properties:

- Relative rarity in the languages of the world.
- Late 'acquisition' by children (typically referring to the recognizability of a form in child speech).
- Loss in aphasia (typically referring to the recognizability of a form in aphasic speech).
- Relative difficulty of perception (not always experimentally validated).
- Relative difficulty of articulation (based on impressions of what is hard to say).
- Tendency to be lost in language change and to not arise in language change.

All of these criteria have been criticized by Hale and Reiss (2000a,b; see references therein). These works conclude that the best way to gain an understanding of the computational system of phonology is to assume that the phonetic substance (say, the spectral properties of sound waves, or the physiology of articulation) that leads to the construction of phonological entities (say, feature matrices) *never* directly determines how the phonological entities are treated by the computational system. The computational system treats features as arbitrary symbols.

An additional consideration in evaluating markedness and the notion of grammatical ill-formedness, pointed out by John Ohala over the years, is that building into a theory of universal grammar the articulatory or perceptual basis for a recurring sound pattern leads to an unacceptable duplication of explanatory mechanisms. For example, if we can show in a laboratory that consonant place distinctions are less salient in codas than in onsets, we do not need to build into phonological theory innate constraints that lead to place feature neutralization in codas. The facts of perception demonstrated in the laboratory, combined with a theory of language acquisition and change, will account for the attested patterns. Offering an additional cause, within the phonology, is just bad science.

Based on learnability considerations, the weakness of markedness considerations, and various conceptual arguments, we conclude that a rule-based derivational model of grammar is superior, since it can be stated in purely positive terms, without prohibitions. The problem of inductive uncertainty is avoided, by the learner and the linguist, if we adopt a rule-based model of grammar, as defined here.

References

Chomsky, Noam. 1986. Knowledge of language. Westport, CT: Praeger.

Epstein, S., E. Groat, R. Kawashima, and H. Kitahara. 1998. A derivational approach to syntactic relations. Oxford: Oxford University Press.

Hale, M., and C. Reiss. 2000a. Substance abuse and dysfunctionalism: current trends in phonology. Linguistic Inquiry 31. 157–69.

Hale, M., and C. Reiss. 2000b. Phonology as cognition. Phonological knowledge, ed. by N. Burton-Roberts, Philip Carr, and Gerry Docherty. Oxford: Oxford University Press.

Karttunen, Lauri. 1993. Finite state constraints. The last phonological rule, ed. by John Goldsmith. Chicago, IL: University of Chicago Press.

Mohanan, K. P. 2000. The theoretical substance of optimality formalism. The Linguistic Review 17. 143–66.

Ohala, J. 1990. The phonetics and phonology of aspects of assimilation. Papers in laboratory phonology I: between the grammar and physics of speech, ed. by J. Kingston and M. Beckman. Cambridge: Cambridge University Press.

Reiss, C. 2003. Deriving the feature-filling/feature-changing contrast: an application to Hungarian vowel harmony. Linguistic Inquiry 34. 199–224.

———. Constraining acquisition without constraints. Rules and constraints in phonology, ed. by A. Nevins and B. Vaux. Oxford: Oxford University Press, to appear.

CHARLES REISS

Russell, Bertrand Arthur William

A leading British philosopher, logician, essayist, pacifist, and political activist, Bertrand Russell was born in 1872 into a liberal and aristocratic family but was sadly orphaned at the age of 4. He grew up as an *enfant terrible* and later a tireless critic of the establishment, only to be invested with the Order of Merit in 1949 and, a year later, awarded the Nobel Prize for Literature. By the time of his death in 1970 at the age of 97, he had authored some 70 books, and around 4,300 articles, book chapters, reviews, pamphlets etc.,

besides more than 61,000 letters and other documents not originally intended for publication. Along with G.E. Moore, he is widely acclaimed as one of the founders of analytic philosophy and, with Kurt Gödel, he shares the reputation of having made the most significant contribution to the study of logic in the twentieth century. An unflinching pacifist, he was imprisoned for involvement in antiwar protests and dismissed from his job at Cambridge University in 1916, and in 1940 from his job at the City College, New York; years later, in 1961, he was again arrested and imprisoned for a week, this time for engaging in antinuclear protests. As an essayist and pamphleteer, Russell addressed a number of issues of contemporary social, political, and moral concern and never shied away from expressing his views boldly and openly as in his book *Why I am not a Christian*.

Russell was a strong advocate of logicism—the view that all of mathematics is ultimately reducible to logic. To him goes the credit for having directed the attention of fellow philosophers and logicians to the utmost importance of detecting and rooting out contradictions in set theory and elsewhere. In 1901, Russell discovered what is famously known in the literature as 'Russell's paradox', later fully elaborated in his work *Principles of mathematics* (1903). The paradox, or antinomy as it is sometimes called, was identified by imagining a set of all sets that were not members of themselves and asking, apropos of that resultant set, if it was a member of itself, the answer being that it can only be a member of itself by not being a member of itself—which is a blatant contradiction. The discovery of this paradox stimulated intense research in set theory and led to several suggested solutions. Russell's own solution was elaborated in his theory of types, which basically consisted in postulating a hierarchy of sentences with strict prohibition on mixing levels and thus preempting the very question that led to the discovery of the paradox in the first place.

Among Russell's most significant contributions to the philosophy of language is his famous 'theory of definite descriptions' put forward in response to the German logician Gottlob Frege's theory of 'sense' (*Sinn*) and 'reference' (*Bedeutung*). Frege had argued that two expressions, say, 'the morning star' and 'the evening star', although clearly different in their senses, can nevertheless have the same referent, namely, in this particular instance, the planet Venus. Despite its intuitive appeal, Russell found the distinction unsatisfactory and was appalled by what Frege had shown to be an important consequence of the distinction for the very enterprise of logic as it had been traditionally conceived. For, in Frege's view, the sentence 'The present king of France is bald', uttered at a time when France was no longer a monarchy (or, equivalently,

when there was no referent to the descriptive phrase 'the present king of France'), was bereft of any definite truth value, because the existence of the referent is a precondition (or, technically, presupposition) for what is asserted by the sentence to be capable of being either true or false. When the presupposition is not true, what is asserted by the sentence will be neither true nor false. This meant that the time-honored principle of logic known as the law of the excluded middle according to which, given a sentence and its (contradictory) negation, one of them has to be necessarily true no matter what the world is like would be, from now on, subject to the additional constraint that stipulated 'provided the sentence's presuppositions are true'. Russell realized that such a consequence was simply devastating for logic's reputation as a 'science of pure reason'.

Russell's alternative analysis began by showing that the logical form of the sentence under discussion was much more complex than Frege had thought: it was, he argued, actually a conjunction of three separate propositions, namely: (1) There is a present King of France. (2) There is only one king of France. (3) Whosoever is the King of France is bald. Now, it is commonplace in logic that the negation of a compound sentence could be shown to be true if any one (or more) of its constituents is known to be false. So, the nonexistence of a King of France was as strong a reason for a sentence affirming his baldness to be false as the fact of the king being hirsute. What Frege took to be a presupposition of the sentence was thus shown to be capable of being treated as one of its straightforward entailments. Whereas the Fregean analysis demanded a thorough revision of one of the fundamental claims on behalf of logic (in addition to the familiar 'true' and 'false', there was the need to admit of 'neither true nor false' as a third possibility), Russell's solution required no new apparatus other than the standard bivalent logic.

From the 1970s onward, many linguists such as Deirdre Wilson (1975) and Ruth Kempson (1975) have followed Russell's lead and urged that many of the phenomena treated by earlier researchers as presuppositions should be better treated as semantic entailments, and any residual problems such as favored readings (including the intuitive feeling that certain presuppositions do hold good in determinate circumstances) should be accounted for by a separate pragmatic component that will have recourse to such theoretical apparatuses as Grice's theory of conversational maxims. The decision to consign part of the problematic to pragmatics may be seen as having as its ultimate inspirational source a key element in P.F. Strawson's objections to Russell's analysis on the grounds that it did not do full justice to ordinary language. (In a protracted polemic with Russell, Strawson

insisted that there was a need to distinguish a sentence or an expression from a use and an utterance thereof.)

Russell's theory of descriptions provided the logical basis for his epistemology centered on a distinction between 'knowledge by acquaintance' and 'knowledge by description'. Among the objects that could be known through acquaintance Russell included what he referred to as 'logically proper names', whose function was to logically refer to certain objects such as sense data and universals. Ordinary names, by contrast, were regarded by Russell as functioning as definite descriptions in disguise. Russell thus pioneered the program known as 'logical atomism', the *locus classicus* of which is Wittgenstein's *Tratactus logico-philosophicus* (translated into English under the same title in 1922), originally prepared under Russell's own supervision. In Russell's view, the world is wholly made up of atomic facts and every meaningful proposition must be composed of constituents with which we are acquainted.

Among Russell's major publications is the monumental work *Principia mathematica* (1910–1913), written in collaboration with his Cambridge tutor Alfred North Whitehead. His 1912 work, *The problems of philosophy*, has long served as one of the popular introductions to the subject for undergraduates.

Biography

Bertrand Russell was born on May 18, 1872 at Ravenscroft, Wales. He was orphaned at the age of 4, with the death of his mother and sister in 1874, and his father in 1876. In 1876, Russell's grandfather, Lord John Russell (a former Prime Minister of Great Britain), and grandmother succeeded in overturning his father's will to win custody of Russell and his brother. With the death of his grandfather in 1878, Russell's grandmother began supervising his upbringing. In 1890, he went to Trinity College, Cambridge, and three years later was awarded a first class B.A. in Mathematics. In 1901, he discovered Russell's para-

dox. From 1907 to 1910 he worked on *Principia mathematica* in collaboration with A.N. Whitehead. He was elected Fellow of the Royal Society in 1908. He was dismissed from Trinity College for antiwar protests (1916), and was imprisoned for six months for involvement in antiwar protests (1918). He was awarded the Order of Merit (1949) and Nobel Prize for Literature (1950). In 1955, he released the Russell–Einstein Manifesto condemning the nuclear arms race. In 1958, he became founding President of the Campaign for Nuclear Disarmament and was imprisoned for one week in 1961 for leading antinuclear protests. He died in Penrhyndeudraeth, Wales, on February 2, 1970.

References

Ayer, A.J. 1972. Russell. London: Fontana/Collins.

Clark, Ronald. 1975. The life of Bertrand Russell. London: J. Cape.

Kempson, Ruth. 1975. Presupposition and the delimitation of semantics. Cambridge: Cambridge University Press.

Irvine, A.D. (ed.) 1998. Bertrand Russell: critical assessments, 4 vols. London: Routledge.

Monk, Ray. 1996. Bertrand Russell: the spirit of solitude. London: J. Cape.

Russell, Bertrand Arthur William, and Alfred North Whitehead. 1910, 1912, 1913. Principia mathematica, 3 vols. Cambridge: Cambridge University Press.

Russell, Bertrand Arthur William. 1912. The problems of philosophy. London: Williams and Norgate; New York: Henry Holt and Company.

———. 1914. Our knowledge of the external world. Chicago and London: The Open Court Publishing Company.

———. 1927. Why i am not a Christian, London: Watts; New York: The Truth Seeker Company.

———. 1929. Marriage and morals. London: George Allen & Unwin; New York: Horace Liveright.

———. 1945. A History of western philosophy. New York: Simon and Schuster; London: George Allen & Unwin, 1946.

———. 1959. My philosophical development. London: George Allen & Unwin; New York: Simon and Schuster.

Sainsbury, R.M. 1979. Russell. London: Routledge.

Wilson, Deirdre. 1975. Presuppositions and non-truth conditional semantics. London: Academic Press.

KANAVILLIL RAJAGOPALAN

Russian and East Slavic Languages

The East Slavic languages—Russian (R), Ukrainian (U), Belarusian (Br)—constitute one of the three branches of the Slavic language family. All three East Slavic languages are written in the Cyrillic alphabet, but Belarusian has a parallel Latin orthographical tradition dating from the mid-sixteenth century.

The East Slavic languages developed from the eastern dialects of Late Common Slavic, when speakers migrated northward and eastward from the presumed Slavic homeland north of the Carpathians around 500 CE into territory sparsely populated by Finno-Ugrian and Baltic peoples. By the eighth century, a lively

trade in furs, amber, wax, honey, slaves, and silver across this vast territory attracted Scandinavian warrior-traders, the Rus', who sought to control the routes that linked the markets of Europe and Asia.

The Rus' dominated the East Slavic tribes in the ninth and tenth centuries, consolidating power first in the lands of the north around Lake Il'men' and at Novgorod, and then in the south at Kiev. They gradually assimilated into the larger Slavic population, leaving traces in the form of a few dozen loanwords (e.g. R *jákor*, U *jákir*, Br *jákar* 'anchor'; cf. Old Swedish *ankari*), but most importantly in the very name Rus', ascribed first to the East Slavs around Kiev and deriva- tively to the Kievan lands. The name ultimately came to refer to East Slavic lands collectively and by exten- sion to all the early East Slavs and to their language, which for the early historical period (up to the four- teenth century) can be called Rusian.

Our direct knowledge of Rusian is complicated by the fact that Rusian Church Slavonic, and not Rusian, was the first written language, gradually introduced after the conversion of the East Slavs to Christianity in 988. Rusian Church Slavonic (hereafter Slavonic) was a local variant of Old Church Slavonic, the liturgical language devised by Saints Cyril and Methodius on the basis of their ninth-century South Slavic (Macedonian) dialect for use in spreading Byzantine Christianity among the Slavs.

Slavonic was basically comprehensible to the East Slavs but contained a number of superficially South Slavic features, primarily of a phonological and mor- phological nature. For example, the Common Slavic roots **gard-* 'fortified town, wall' and **berg-* 'bank, shore' were realized as *grad-*, *brêg-* in Slavonic, where- as the vernacular Rusian had the regular East Slavic pleophonic counterparts *gorod-*, *bereg-*, respectively. Similarly, the Slavonic and Rusian reflexes of the Common Slavic sequence **tj* were *šč* (*shch*) and *č* (*ch*), respectively, and thus Slavonic *svêšča* and Rusian *svêča* 'lamp; candle', both derived from the Common Slavic root **svêt-* 'light' plus the suffix *-j-*. Despite such formal differences, the basic grammar and the majority of the most common lexical items in Slavonic and Rusian were identical. Some scholars prefer to treat them as variants or registers of a single language, whereas others view them as two closely related but separate languages. The latter view is adopted here.

In addition to the superficial distinctions noted above, written Slavonic differed from vernacular Rusian in its richer lexicon, and in its potentially more complex syntax, both inspired by elements of the Bulgarian Slavonic, Serbian Slavonic, or Greek texts that served as sources for translation. Although the pri- mary written record of early Rus' appeared in Slavonic rather than Rusian, lapses in the copying and editing of texts provide evidence of the actual speech habits of the Rusian scribe producing them.

Since the 1950s, archeological exploration has revealed direct examples of written Rusian in docu- ments uncovered in predominately northwestern locales, especially Novgorod, Pskov, and Staraja Russa. Unlike their Slavonic counterparts preserved on costly parchment, these legal and financial docu- ments or personal correspondence were scratched onto disposable birchbark. They provide valuable examples of the Rusian vernacular in a geographically circum- scribed area from the eleventh through the fifteenth centuries.

From the very beginning, strictly Rusian features mingled with those of Slavonic, a situation that became more common as new texts, created and borrowed, provided fresh opportunities for linguistic choice. Generic and thematic factors certainly influenced the predilection for Slavonic in religious and other solemn texts and Russian in mundane ones, but they were not overriding. At times, for example, a scribe might pre- fer Slavonic *gradъ* over Rusian *gorodъ* simply to save the space of one letter on the expensive parchment at the end of a line.

Although there was dialectal divergence among the East Slavs by the time of Christianization, virtually no evidence exists to suggest that it caused any difficulty in communication between residents of the Novgorodian north and the Kievan south. For this reason, the vernac- ular language of the early East Slavs is best considered as Rusian alone, with dialectal variants distributed geo- graphically. Nonetheless, linguistic evolution and his- torical circumstance helped to extend and deepen these distinctions to a point that it is more appropriate to speak of separate languages rather than dialects.

The effects of the phonological change known as the 'jer shift' produced dramatic differences north and south, alternative choices that affected consonant and vowel inventory as well as morphophonemic and mor- phological alternation (see below). Historical and political factors also served to create distinct commu- nities of speakers whose speech patterns became increasingly divergent from one another and subject to influence from other languages. For example, the growing importance of such northeastern centers as Rostov, Suzdal', and Vladimir (twelfth century) in the face of older Novgorod and Kiev; the Mongol inva- sions (mid-thirteenth century); the Novgorodian colo- nization of the north (thirteenth to mid-fifteenth centuries); the incorporation of southwestern Rus' into the Grand Duchy of Lithuania (thirteenth to fourteenth centuries) and Poland (fourteenth century); and the ascendancy of Moscow (fourteenth to fifteenth cen- turies) all contributed to the enhancement of regional linguistic difference.

By the fourteenth century, two language areas of Rus' had developed: a northern, Russian one destined to come under the hegemony of Moscow, and a southern, Ruthenian one, at first found largely within the Lithuanian political orbit, and then from 1569, under the rule of the Polish–Lithuanian Commonwealth. In Ruthenian territory, the prominent distinctions between Belarusian and Ukrainian were apparent by the fifteenth to sixteenth centuries, although their beginnings are discernable three to four centuries earlier. From the mid-seventeenth century, parts of the Ukrainian and Belarusian territories gradually fell under the control of Muscovy, then the Russian Empire, and finally the Soviet Union, a process that extended well into the twentieth century.

Consideration of the structure and dialectal distribution of Rusian provides a common basis for comparing and contrasting the development of the East Slavic languages with respect to each other. Rusian had a rich system of vowels (*i,y,u,ê,e,o,ä,a,ь,ъ*), including two high, lax vowels, the jers (symbolized as front *ь* and back *ъ*, respectively), a higher-mid tense front vowel (symbolized as *ê*), and a low tense front vowel (symbolized as *ä*). Over the course of the eleventh and twelfth centuries, the jers were either reidentified as mid-vowels *e* and *o*, or were lost (e.g. Russian *pьnь* 'stump', *sъnъ* 'sleep; dream' > R,Br *p'en'*, *son*, U *pen'*, *son*). The jer shift reduced the number of distinctive vowels by a third and nearly doubled the inventory of consonants by creating sets of distinctively paired plain (hard) labial and dental consonants (*p,b,m,v,t,d,s,z,n,l,r*) and palatalized (soft) counterparts (*p',b',m',v',t',d',s',z',n',l',r'*). Russian has largely preserved this later Rusian phonemic inheritance, whereas Ukrainian, apart from the dispalatalization (hardening) of consonants before Rusian *i* and *e*, preserved most soft dentals, but hardened soft labials. Belarussian preserved soft labials before vowels, hardened *r'* to *r*, and affricated *t'* and *d'* to *c'* (*ts'*) and *ʒ'* (*dz'*), the result of assibilation, so-called *cekan'e* and *dzekan'e*, respectively. After the jer shift, most new sequences of dentals or palatals plus *j* yielded soft geminate (doubled) dentals and palatals (*r'* and *j* are the exceptions) in Ukrainian and Belarusian, a reflex not typical of Russian, e.g. Rusian *žitьje* 'life', U *žyt't'a*, Br *žic'c'o*, R *žit'jo*.

Many Rusian dialects had a seven-vowel system after the jer shift: *i,e,a,o,u* plus the tense vowels *ê* and *ô*. The latter two were distributed differently, north and south, and were ultimately the most unstable. At the present time, standard Russian and Belarusian each have five-vowel systems (*i,e,a,o,u*), whereas standard Ukrainian has a six-vowel system (*i,y,e,a,o,u*). The higher-mid front *y* is a characteristically Ukrainian distinctive vowel, derived from merged Rusian *i* and *y*,

with *ê* (and typically *ô* as well) subsequently raising and fronting to *i* (so-called *ikavism*), e.g. Russian *lêsъ* 'forest', *stolъ* 'table' > U *l'is*, *st'il*.

The oldest layers of dialectal differentiation in Rusian follow a path from southwest to north and northeast, reflecting the early direction of Slavic development from the original Carpathian center to the northern and eastern periphery. Later innovation is stimulated by the rise of new centers of cultural diffusion, as noted above. Thus, among the oldest isoglosses is one that distinguishes dialects that preserve Common Slavic *g as a stop (the more 'peripheral' North Russian) from those that have lost closure, producing a fricative like γ or *h* (the more 'central' South Russian, Belarusian, Ukrainian). Another isogloss distinguishes the development of the tense jers (in the environment before *j*) as mid-vowels *e* and *o* (the more 'peripheral' Russian) or vowels *i* or *y* (the more 'central' Ruthenian), e.g. Rusian *mъju*, R *moju*, U *myju*, Br *miju* [mýju] 'I wash'. A third, involving the hardening of palatalized labial and dental consonants before Rusian *e*, separates the dialects of Russian and Belarusian (presence of palatalization) from those of Ukrainian (absence of palatalization). Some archaic Carpathian dialects of Ukrainian have preserved the Rusian phonemic distinction between *i* and *y*, unlike the rest of East Slavic.

Rusian, like Slavonic, had a rich system of grammatical categories. Nouns, pronouns, and adjectives were inflected by gender (masculine, feminine, neuter), number (singular, dual, plural), case (nominative, accusative, genitive, locative, dative, instrumental), animacy, and a special vocative form. Verbs were capable of distinguishing mood, tense, voice, and aspect. Participles and adjectives were inflected by gender, number, and case according to two declensions, a 'short' nominal declension and a 'long' adjectival declension, the latter providing expression for definiteness or generic status. All three East Slavic languages have retained the dynamic, free stress characteristic of Rusian.

By the fourteenth century, a number of Rusian grammatical categories had been lost, no longer productive in any of the East Slavic languages. These include dual number, the aorist and imperfect tenses, and case in short adjectives and participles. The perfect took over the function of past tense in all of East Slavic. Ukrainian has preserved a separate vocative form. Ukrainian and Belarusian both have a rarely encountered pluperfect tense. The morphophonemic velar ~ dental sibilant alternation found in most Rusian dialects was retained on a more limited basis in Ukrainian and Belarusian, but lost in Russian (e.g. Rusian *ruka/rucê* 'hand, arm' nom.sg./dat.-loc. sg., U *ruka/ruc'i*, Br *ruka/ruce*, but R *ruka/ruk'e*). Of syntactic interest are

the many East Slavic constructions without overt nominative subjects (e.g. R *mn'e ujt'i* 'I have to go' [lit. to me to go], U *cerkvu bulo zbudovano* 'The church was built' [lit. church~acc.~ was~neut.sg.~ built~neut.sg.~], Br *m'an'e n'e bilo* [byló] *doma* 'I wasn't home' [lit. of me not was-~neut.sg.~ at home]).

Russian

Spoken by approximately 275 million people, 165 million of them natively, Russian is by far the largest of the Slavic languages. The vast majority of Russian speakers reside in the Russian Federation or constitute substantial minorities in several of the republics of the former Soviet Union, including Ukraine, Belarus, Latvia, Kyrgyzstan, and Kazakhstan. It is the major language of the Russian Federation and one of the official languages of Belarus and Kazakhstan. There are significant Russian-speaking communities in Western Europe, the United States, and Canada. Russian is one of the five official languages of the United Nations.

The history of written language in Russian territory is one of continuing tension between Slavonic and vernacular Russian from the fourteenth to the eighteenth centuries. The church-supported emergence of Moscow as the center of the evolving Russian state saw the development of Slavonic into an ever more ornamental, complicated medium with intentional archaization from the late fourteenth century onward, a trend that resonated with the late fourteenth-century reforms in Bulgaria. These changes had the effect of distancing Slavonic from the spoken language by eliminating vernacular forms and introducing artificial elements and grammatical patterns. At the same time, during the fifteenth and sixteenth centuries, the language of the growing Muscovite bureaucracy—the formulaic, vernacular-based chancery language—displayed a normative tendency that elevated the speech patterns of Moscow and avoided the dialectal features of other areas of Russia.

The union of Left-Bank Ukraine and Muscovite Rus' in the mid-seventeenth century provided Ukrainian bookmen with the opportunity to bring to Muscovy their better developed grammatical tradition based on Latin and Greek models. They codified and enriched the Slavonic tradition, particularly with foreign vocabulary from the West, including Latin, primarily through Polish and Ukrainian intermediaries. In their turn, Slavonic and the chancery language influenced the development of vernacular Russian. It was only in the mid-eighteenth century, especially through the grammatical and lexical reforms of Adodurov and especially Lomonosov, that there was a reasoned attempt to control the distribution of foreign, Slavonic, and Russian elements in the Russian written language.

Karamzin in the late eighteenth to early nineteenth century and then Pushkin helped to refine the Russian language by example, setting the stage for it to become a vehicle capable of artistic and scientific expression on the highest levels of European culture.

Russian is traditionally divided into three major dialect zones: north, south, and central. South Russian is distinguished by the presence of two innovations: (1) the fricative γ instead of original stop *g* and (2) *akan'e*. Akan'e refers to the pronunciation of unstressed *o* and *a* as [a] initially or after a hard consonant in the first pretonic syllable (the syllable immediately preceding the stressed one) and as central lax schwa elsewhere (symbolized [ə]): compare *molodój* [məladój] 'young' and *nagradnój* [nəgradnój] 'pertaining to reward'. The large transitional zone between North and South Russian is composed of the central dialects (including Moscow), which show their intermediate status in favoring northern consonantism with *g* and southern vocalism with akan'e. It is the Muscovite dialect that ultimately served as the basis of the standard language: compare NoR *mogú* [mogú], SoR *moγú* [maγú], and standard Russian *mogú* [magú]. In addition to akan'e, standard Russian is also characterized by *ikan'e*, the pronunciation of unstressed *o,a,e* as [i] in the first pretonic syllable after a soft consonant, e.g. *n'os'í* [n'is'í] 'carry' (imperative 2sg.), *zan'alá* [zən'ilá] 'occupy' (past fem.sg.), *n'emój* [n'imój] 'dumb' (adjective nom.sg.masc.)

Ukrainian

Ukrainian is spoken by approximately 49 million people, making it the second largest Slavic language. Most of the speakers of Ukrainian reside in Ukraine, but there are substantial communities in Russia, Poland, Slovakia, Canada, and the United States. Ukrainian is the state language of Ukraine. In addition to Ruthenian (*rus'ka mova*) and plain talk (*prosta mova*), it was also called Little Russian from the mid-seventeenth century to the early twentieth, a derivative from Little Russia, itself a calque of a Byzantine Greek administrative term used to distinguish the more proximate Ukrainian territory from the more distant Muscovite (called Great Russia).

From the late fourteenth century, Ruthenian Slavonic in Ukrainian and Belarusian territories underwent archaization analogous to that of Slavonic in Russian territory. By this time, Galicia had been absorbed into the Kingdom of Poland and the rest of the Ukrainian and Belarusian territories into Lithuania. This turn of events had far-reaching linguistic implications for both Ruthenian and Slavonic. In Polish Galicia the vernacular-based Ruthenian chancery language was ultimately replaced by Latin in accordance with a 1433 decree. In

Lithuania, on the contrary, Ruthenian was elevated in status to become the official language of government administration. Based on the East Slavic vernacular spoken around Vilna (Vilnius), the chancery language developed as a kind of koine, which favored specific Belarusian elements over Ukrainian ones, but avoided the more remarkable characteristics of either, such as Belarusian cekan'e/dzekan'e and akan'e, and Ukrainian ikavism, in its various developmental stages. By the middle of the sixteenth century, spoken and written Ruthenian was making its influence felt in Slavonic in the presence of greater numbers of vernacular forms.

The 1569 Union of Lublin resulted in the merger of Poland and Lithuania as the Polish–Lithuanian Commonwealth. With Poland as the dominant partner, Polish and, to a lesser extent, Latin were introduced as the official languages of government administration throughout the land. Without administrative function, Ruthenian became subject to local adaptation. Features from the western and northern Ukrainian dialects began to dislodge their Belarusian counterparts in the Ukrainian territories to the south, whereas more prominent Belarusian features were developed in the Belarusian territories to the north. Both were subject to heavy Polonization and influence from Slavonic as well.

After the 1596 Union of Brest brought together Catholics and Orthodox Uniates, Polish and Latin began to challenge Slavonic as well in the spiritual sphere. A Slavonic revival begun in the western Ukrainian town of Ostrih resulted in the publication of a complete translation of the Bible in 1581. The revival spread later to L'viv and Kyiv and resulted in new Slavonic dictionaries and grammars.

Ukrainian had become the administrative language in Ukrainian territory east of the Dnieper after an uprising in 1648 resulted in the establishment of a Cossack state, the Hetmanate. It retained that role even after the union with Muscovy in 1654. Mazepa's unsuccessful attempt in 1709 to wrest the Hetmanate from Russian control led to the replacement of Ruthenian Slavonic by Russian Slavonic. Late in the century, Russian replaced Ukrainian as the language of administration, both in the territories of the old Hetmanate, and in the Right-Bank Ukrainian territory acquired in the 1793 and 1795 partitions of Poland.

The emergence of the modern Ukrainian standard language begins with Kotljarevs'kyj's translation-reworking of the *Aeneid* (1798, 1809, 1842), in which he showed a preference for southeastern dialect features (Poltava, Kharkiv, Kyiv). He prepared the way for a flowering of literary Ukrainian during the period of Ukrainian Romanticism (1820s–1840s), chiefly found in the writings of Kvitka, Ševčenko, and Kuliš, all of whom sought to realize a synthesis of Ukrainian

and Slavonic, although not to the degree seen in Russian. Harsh restrictions were imposed by the Russian government on the use of Ukrainian in 1863, leading to the ban on its public use in 1876, a prohibition lifted only in 1905. Meanwhile, the Ukrainian literary tradition that had continued in Polish Galicia (ceded to Austria in 1772, 1795) had an impact on the future development of Ukrainian in eastern Ukraine. The two centers of Ukrainian culture were joined only after World War II. The Soviet period was marked by an intense government campaign of Russification.

Modern Ukrainian competes with Russian and a Russianized Ukrainian hybrid called *surzhyk*, especially in the capital of Kyiv and in eastern Ukraine. Since independence (1991), Ukrainian, the sole state language, has gained considerable ground in the official sphere, especially in government, communication, and education.

Ukrainian is traditionally divided into three major dialect zones: northern, southwestern (including the Carpathians, eastern Poland, and eastern Slovakia), and southeastern. A typical feature of the northern dialects is the presence of diphthongs from older *ê* and *ô* under stress that correspond to monophthongal vowels in the standard language in either the southwest or southeastern variant. The southeastern dialects are characterized by ikavism, whereas this process has been only partly realized in many southwest dialects.

Belarusian

Belarusian has approximately 10 million speakers, most of whom reside in Belarus. There are small communities of Belarusian speakers in Poland, the United States, and Canada. One of the official languages of Belarus, Belarusian (also rarely Belarusan) has been called Belorussian or Byelorussian, a name influenced by the incorporation of Belarusian territory into the Russian Empire. The English rendering of Belarusian as 'White Russian' produces an unfortunate ambiguity with the political term referring to the enemies of the Bolshevik Revolution in Russia.

Following the Union of Brest in 1596, Belarusian (plain talk) increasingly yielded to Latin and finally to Polish as the authoritative written language, and in 1697 it was banned from administrative use completely. During the eighteenth century, Belarusian was largely limited to the burlesque interludes for religious school dramas, themselves written in Slavonic, Latin, or Polish. After Belarusian territory was incorporated into Russia during the partitions of Poland in 1772 and 1795, the Russian government banned the use of Belarusian in schools and in publications, a proscription that lasted until 1905. The publication of a

normative grammar by Branislaw Taraškevič (1918 and later editions) was a major step toward the establishment of a standard literary language. Belarusian was subjected to heavy Russification during the Soviet period. Its fate remains precarious in contemporary Belarus: even in the capital of Minsk, one is more likely to hear Russian or a Russianized Belarusian hybrid called *triasanka* than Belarusian itself.

There are three major dialect areas of Belarusian: northeast, southwest, and a broad central one (including Minsk), transitional to both. The northeastern dialects, themselves phonologically and morphologically transitional to Russian, feature dissimilative akan'e and jakan'e. *Jakan'e* refers to the neutralization of *o,a,e* in unstressed environments after soft consonants in which one of the pronounced variants is [a]. The dissimilative principle yields two basic variants of neutralization in the first pretonic syllable: (1) the low vowel [a] and (2) a nonlow vowel, commonly [ə] in akan'e and [i] in jakan'e. The first pretonic vowel in each case dissimilates the stressed (tonic) vowel in height: if the tonic vowel is high, the first pretonic variant is low, and vice versa; thus *vodí* [vadý] 'water' (gen.sg.), but *vodá* [vədá] 'water' (nom.sg.), *n'osú* [n'asú] 'carry' (nonpast, 1st sg.), but *n'os'éš* [n'is'éš] 'carry' (nonpast, 2nd sg.). The southwestern dialects are phonologically and morphologically transitional to Ukrainian. In all but the southernmost regions, which are devoid of akan'e and jakan'e, strong (nondissimilative) akan'e and jakan'e are predominant: the first pretonic neutralized variant of *o,a,e* is always realized as [a], regardless of the nature of the tonic vowel; thus [vadý], [vadá], [n'asú], [n'as'éš]. Standard Belarusian, based on the central dialects (including Minsk), has strong akan'e and jakan'e as well.

References

Avanesaw, R.I. 1964. Narysy pa belaruskaj dyjalektalohii. Minsk: Navuka i tèxnika.

Avanesov, R.I., and V.G. Orlova. 1965. Russkaja dialektologija. Moscow: Nauka.

Comrie, Bernard, and Greville G. Corbett (eds.) 1993. The Slavonic languages. London and New York: Routledge.

Issatschenko, Alexander V. 1980. Russian. In Schenker and Stankiewicz.

Mayo, Peter. 1993. Belorussian. In Comrie and Corbett.

McMillin, Arnold. 1980. Belorussian. In Schenker and Stankiewicz.

Schenker, Alexander M., and Edward Stankiewicz (eds.) 1980. The Slavic literary languages: formation and development. New Haven: Yale University Press.

Shevelov, George Y. 1979. A historical phonology of the Ukrainian language. Heidelberg: Carl Winter Universitätsverlag.

———. 1980. Ukrainian. In Schenker and Stankiewicz.

———. 1993. Ukrainian. In Comrie and Corbett.

Strumins'kyj, Bohdan. 1984. The language question in the Ukrainian lands before the nineteenth century. Aspects of the Slavic language question, Vol. 2: East Slavic, ed. by Riccardo Picchio and Harvey Goldblatt. New Haven: Yale Concilium on International and Area Studies.

Timberlake, Alan. 1993. Russian. In Comrie and Corbett.

Vlasto, A.P. 1986. A linguistic history of Russia to the end of the eighteenth century. Oxford: Clarendon Press.

Wexler, Paul. 1977. A historical phonology of the Belorussian language. Heidelberg: Carl Winter Universitätsverlag.

Worth, Dean S. 1998. Language. Modern Russian culture, ed. by Nicholas Rzhevsky. Cambridge: Cambridge University Press.

Žylko, F.T. 1966. Narysy z dialektolohiji ukrajins'koji movy, 2nd edition. Kyiv: Radjans'ka Škola.

MICHAEL S. FLIER

S

Sandhi

Sandhi rules are phonological alternations that are triggered at junctures of words or morphemes. The name sandhi comes from Sanskrit and means 'juncture'. It does not designate one particular phonological process. Instead, it is a nonspecific cover term for any kind of sound mutation that occurs at the edges of words and morphemes, and that is triggered in environments created by morphological or syntactic concatenation operations. Sandhi phenomena are very common across the languages of the world. Some of the best-known examples are tone sandhi, which is found in Chinese for example, Welsh and Irish consonant mutations, and a set of Sanskrit sandhi rules.

Two types of sandhi can be distinguished: internal and external sandhi. Internal sandhi rules are triggered word internally, between two morphemes. External sandhi rules operate at word edges, when two words become adjacent as a result of some syntactic process in the language.

An example of internal sandhi can be found in English words such as symbol and symbiosis. Both words are derived by addition of the prefix *syn* to the stem. In its basic form, this prefix can be observed in words like synthesis, synopsis, syntax, synchronicity, synonym, etc. When combined with a stem beginning with /b/, as in the words symbol and symbiosis, the last consonant of the prefix changes to /m/. This internal sandhi rule is an instance of the phonological process known as place assimilation, in which the place of articulation of some consonant becomes dependent on the place of articulation of an adjacent consonant. In the case at hand, the place of articulation of the consonant /n/ becomes identical to the place of

articulation of the consonant that follows it. The latter consonant can therefore be said to act as the trigger for this phonological change.

One of the internal sandhi rules of Sanskrit is a phonological process known as voice assimilation. The phonological process in question leads to a change of voiceless consonants, such as /p/, /t/, and /d/, turning them into their voiced counterparts /b/, /d/, and /g/. The effect of this rule is observable in the compound *sadaha* 'good day' derived from *sat* 'good' and *aha* 'day'. Compound formation leads to a configuration in which the two segments—the devoiced word final /t/ and the voiced, word initial /a/—become adjacent, and thus creates the conditions for the application of voice assimilation.

Interestingly, the same rule that operates word internally in Sanskrit, between two portions of a compound, also applies as external sandhi between two words of a phrase. For instance, the phrase *samyag uktam* 'correctly spoken' consists of the word *samyak* 'correctly' and *uktam* 'spoken'. Again, voice assimilation of the word-final consonant, in this case /k./, can be observed under the influence of the following voiced segment.

In English, external sandhi may occur in phrases such as *miss you, kiss you, caught you,* etc, where the word final consonant of the verb changes to [sh] or [ch] to give *mi[sh] you, ki[sh] you,* and *cough[ch] you.* This phonological process is known as *palatalization.* Another example of external sandhi is known by the name of *intrusive-R.* It exists in some, but not all, dialects of English. It is a feature of the nonrhotic variants (i.e. those in which the word final /r/ as in *bar, star, Zanzibar,* etc. is not pronounced). Intrusive-R is an

insertion of [r] between two words of which the first word ends in a vowel, while the second begins with one, e.g. *vanilla-r-ice-cream, the idea-r-of it, I saw-r-it,* etc.

An external sandhi rule that is present in some dialects of Italian is known as raddoppiamento sintattico. It is triggered only when the two words between which the sandhi operates are contained in the same local domain. The phonological process involved is the lengthening (or doubling) of the word-initial consonant. It is triggered when the word is preceded by another word ending with a stressed vowel, eg. *città **pp**ulita* 'clean city' from *città* 'city' and ***p**ulita* 'clean', or *metà **ll**ibro* 'half book' from *metà* 'half' and *libro* 'book'. However, for consonant doubling to take place, the words have to be contained within the same phonological phrase. When this requirement is met, as in (1a), and the two words are found within the relevant domain (i.e. the phonological phrase here marked by the subscript Φ), raddoppiamento sintattico takes place. On the other hand, if the two words are separated by a phonological boundary (marked as //), raddoppiamento sintattico fails. This is illustrated by (1b), where the word-initial consonant in *molto* 'old' fails to undergo lengthening.

(1a) (Avra **ll**etò)$_{\Phi}$ (il libro)$_{\Phi}$.
       'He will have read the book.'
(1b) (Visita la citt*à*)$_{\Phi}$ // (molto vechie)$_{\Phi}$.
       'He visits very old cities.'

An important feature of external sandhi is that an application of the phonological process in question is determined by the phrasal groupings of the words involved, as shown by the data in (1a) and (1b). Since the phrasal organization of the sentence is determined by syntax, the conditions under which external sandhi rules apply are to some extent influenced by this component of the grammar. However, this influence only goes some way. The constituents into which words are syntactically organized such as syntactic phrases are not isomorphic to the groupings of words into phonological constituents such as phonological phrases, although there are certain connections between the two. The relationship of syntax and phonology (the syntax–phonology interface) and the field of prosodic phonology in general is a large and fertile field of research (Nespor and Vogel 1986; Selkirk 1984).

References

Andersen, Henning (ed.) 1986. Sandhi phenomena in the languages of Europe. Berlin, New York and Amsterdam: Mouton de Gruyter.

Chen, Matthew Y. 2000. Tone Sandhi: patterns across Chinese dialects. Cambridge: Cambridge University Press.

Inkelas, Sharon, and Draga Zec. The phonology syntax connection. Chicago and London: University of Chicago Press.

Nespor, Marina, and Irene Vogel. 1982. 'Prosodic domains of external Sandhi rules' The structure of phonological representations, Part I, ed. by Harry van der Hulst and Norval Smith, 225–55. Dordrecht: Foris Publications.

———. 1986. Prosodic phonology. Dordrecht: Foris Publications.

Selkirk, Elisabeth O. 1984. Phonology and syntax: the relation between sound and structure. Cambridge, MA: MIT Press.

AMELA CAMDZIC

Sanskrit

Sanskrit belongs to the Indo-Iranian branch of Indo-European, whose subgroupings are Iranian and Indo-Aryan. Indo-Aryan includes Vedic Sanskrit, Classical Sanskrit, and the Prakrits, or 'vulgar' varieties, which have as their offshoots the modern Indo-Aryan languages, such as Hindi-Urdu, Bengali, Gujerati, Marathi, Panjabi, Sinhalese, and Romani.

Sanskrit is one of the most anciently attested Indo-European languages and has played a crucial role in the reconstruction of Proto-Indo-European. Apart from some isolated forms (see below), the most ancient literary sources are the poems collected in the books of *Veda* (hence the name 'Vedic Sanskrit'), which record an oral tradition, going back to the second millennium BCE. Classical Sanskrit is a highly standardized language, as such used without interruption from its first written attestations, and virtually unchanged until the present; however, in the early phases dialectal variation is visible, even between Vedic and Classical Sanskrit. The former is not the direct ancestor of Sanskrit: its peculiarities include both archaisms and innovations with respect to the classical literary language, pointing toward some kind of diatopic variation. Furthermore, the Prakrits of the Middle Indian period (300 BCE–200 CE) do not derive directly from Classical Sanskrit, but display a variety of different features, some directly connectable to Vedic. Among the Prakrits, the most important as a

literary language is Pāli, the language of the Buddhist Canon. The coexistence of Sanskrit with Prakrits is ancient, and is mostly a matter of social variation: in classical drama, socially inferior characters (including women) usually speak Prakrit.

The Aryan populations reached India toward the end of the second millennium BCE, and established themselves in the upper Indus valley. Before that time, the Aryans had already built a linguistic group distinct from the Iranians, as shown by the traces left during their journey through the Middle East. In the first half of the second millennium BCE, an Aryan aristocracy dominated the Hurrian kingdom of Mitanni, located on the upper course of Tigris and Euphrates. Names of Indic deities are mentioned in a treaty with a Hittite king, and some Indic words are found as loanwords in a Hurrian text of the period, among which are some numerals, which attest a pre-Sanskrit variety (e.g. *aika*, 'one', in which the monophthongization /ai/ > /e/ had not yet taken place, cf. Sanskrit *eka*).

Grammatical Sketch

Phonology and Script

Sanskrit is written in *devanāgari* script, still used for the modern Aryan varieties, which derives from an earlier script, called *brāhmī*. The *devanāgari* is a peculiar writing system, halfway between a syllabary and an alphabet, consisting of 13 signs for vowels and 35 for consonants, plus a number of diacritics and several compound signs used for representing consonant clusters. Consonant signs not followed by any diacritic are understood as containing the consonant and the vowel /a/.

The *devanāgari* script is very well suited for the representation of Sanskrit phonemes, each sign corresponding to a single phoneme (except for diphthong signs) and each phoneme being represented by a sign. Thus, Table 1 can also serve as an inventory of Sanskrit phonemes.

An important feature of Sanskrit phonology, especially in comparison with the other ancient Indo-European languages, lies in the phonologization of both voiced and voiceless aspirates, the existence of a series of retroflex (also called 'cerebral') stops, borrowed from the non-Indo-European substrate languages of India, and the retention of vocalic liquids, partly with length opposition. (Length opposition is only relevant for /r̥/ ~ /r̥ː/; besides, /l̥/ is a very rare phoneme.) From a diachronic point of view, the most significant innovations are found in the vowel system. At an early stage in Proto-Indo-Iranian, the three vowels of Proto-Indo-European, /a/, /o/, and /e/, merged in /a/ (this change involved both short and long vowels). Later, after the separation of Iranian and Indo-Aryan,

short diphthongs underwent monophthongization: /ai/ > /eː/, /au/ > /oː/. Hence, already in Vedic there is again a distinction, at least for long vowels, but the /eː/'s and /oː/'s of Sanskrit do not reflect the same vowels of Proto-Indo-European.

Sanskrit belongs to the so-called *satəm* group within Indo-European, i.e. it is one of those languages, along with the remaining Indo-Iranian languages, Slavic, and Armenian, in which velars changed into spirants or affricates, and labiovelars changed into velars.

A major role in Sanskrit phonology is played by *sandhi*, i.e. a set of phonological rules determined by the clustering of vowels or consonants.

Morphology

Sanskrit morphology is extremely rich. Inflectional classes of nouns are based on the Indo-European distinction between thematic and athematic declension, with subclasses of the latter for stems in nasal, liquid, /i/, /u/, and long vowel. Nouns have eight cases (nominative, vocative, accusative, instrumental, dative, ablative, genitive, and locative) and three numbers (singular, dual, and plural); gender distinctions include masculine, feminine, and neuter. Adjectives, which also inflect for degrees of comparison (comparative and superlative), follow the inflection of nouns for number and case, while pronouns follow special patterns of inflection.

Verbs are traditionally divided into ten classes (eight in Vedic), of which the first, fourth, sixth, and tenth follow the thematic conjugation, while the remaining follow the athematic conjugation. In the thematic conjugation, the thematic vowel -*a*- (< PIE *-o-) is inserted between the stem and the ending; note that, contrary to the other Indo-European languages, which have -*ō* as ending of the first person singular of the present indicative, Sanskrit has reintroduced the ending -*mi* of the athematic declension, so that we find a cumulation of both endings: *bharāmi*, 'I bring', first class, thematic, built with the present stem *bhar-*, ending-*ā* plus ending -*mi* (cp. with Latin *fer-ō*, same meaning). Verbs distinguish three persons in the singular, dual, and plural.

Tenses include the present, the imperfect, the aorist, the perfect, and the future; in addition, Vedic also had a compound pluperfect. In origin, the verb system was based on an aspectual opposition between present and perfect. The present denoted actions or processes, while the perfect denoted states. The Indo-European opposition between perfective/imperfective is continued only formally, in the opposition between aorist and present, but not semantically (see below). Verbs have a special stem for the present, the perfect, and the aorist; the differences among stems are usually based on root gradation, which opposes a reduced

Devanāgarī alphabet

Primary vowels

	Short				Long				Diphthongs		
	Initial		Diacritic		Initial		Diacritic		Initial		Diacritic
Unrounded low central	अ	a	प	pa	आ	ā	पा	pā			
Unrounded high front	इ	i	पि	pi	ई	ī	पी	pī			
Rounded high back	उ	u	पु	pu	ऊ	ū	पू	pū			
Syllabic variant	ऋ	r̥	पृ	pr̥	ॠ	r̥̄	पॄ	pr̥̄			

Secondary vowels

Unrounded front					ए	e	पे	pe	ऐ	ai	पै	pai
Rounded back					ओ	o	पो	po	औ	au	पौ	pau

Other symbols

अं	aṅ	*anusvāra* - nasalises vowel		अँ	aṃ	*anunāsika/candrabindu* - nasalises vowel
अः	aḥ	*visarga* - adds voiceless breath after vowel		प्	p	*virāma* - mutes vowel

Consonants
Occlusives

	Voiceless plosives				Voiced plosives				Nasals	
	unaspirated		aspirated		unaspirated		aspirated			
Velar	क	ka	ख	kha	ग	ga	घ	gha	ङ	ṅa
Palatal	च	ca	छ	cha	ज	ja	झ	jha	ञ	ña
Retroflex	ट	ṭa	ठ	ṭha	ड	ḍa	ढ	ḍha	ण	ṇa
Dental	त	ta	थ	tha	द	da	ध	dha	न	na
Labial	प	pa	फ	pha	ब	ba	भ	bha	म	ma

Sonorants and fricatives

	Palatal		Retroflex		Dental		Labial	
Sonorants	य	ya	र	ra	ळ	la	व	va
Sibilants	श	śa	ष	ṣa	स	sa		

Other letters   Variant letters used in Mumbai

ह	ha	ऴ	ḷa	झ़	jha	ण	ṇa

A selection of conjunct consonants

क्ष	kṣa	ज्ञ	jña	त्त	tta	त्र	tra	प्य	pya	त्क	tka	ट्क	ṭka	ह्य	hya	त्त्व	ttva

Numerals

०	१	२	३	४	५	६	७	८	९	१०
0	1	2	3	4	5	6	7	8	9	10

(Source: http://www.omniglot.com/writing/devanagari.htm)

grade (e.g. *kṛ-*, 'make') to full grade, also called *guṇa* (*kar-*), to lengthened grade, also called *vṛddhi* (*kār-*). (Note that there is no one-to-one correspondence between vowel grade and a specific tense.) The perfect is usually made on a reduplicated stem (e.g. *ca-kar-a*, 'I have made').

Already in Vedic Sanskrit, the aspectual value of the aorist was given up: the aorist is used in reference to recent past, and never without past reference in contrast to the present (as was the Ancient Greek aorist). The perfect added to its original stative meaning a more recent resultative meaning, which also had its temporal implication of remote past. Furthermore, the so-called imperfect, derived from the present tense, was used in narration, where it did not display any particular tendency toward expressing the imperfective aspect.

In Classical Sanskrit, moods include the indicative, the optative, and the imperative for the present tense; the aorist has an indicative and a precative, and the future has a conditional. Vedic had a more developed mood system, which included an indicative, a subjunctive, an optative, and an injunctive, all made on all three stems (present, aorist, and perfect). From a typological point of view, the most interesting feature of this mood system is constituted by the injunctive. In a way, it stands outside mood oppositions, and is used for general statements.

Voice includes an active (*parasmaipada*) and a middle (*ātmanepada*); not all verbs have both voices, but for those that do, the middle mostly has reflexive value. Furthermore, the middle can have passive meaning, except for the present stem, which builds a special passive by adding the suffix *-ya-*.

The Sanskrit verb has a variety of nominal forms: active and middle present, future and perfect participles, a past passive participle, from which a past active participle is also made, a future passive participle, also called gerundive, an indeclinable participle, also called gerund, and an infinitive.

A peculiarity of Sanskrit is the large extent to which composition is used for building nouns. Following the Indian classification, types of compound include *dvandva* (copulative compounds, e.g. *siṃhagajāḥ*, 'lions and elephants', from *siṃha-*, 'lion', plus *gaja-*, 'elephant'), *tatpuruṣa* (determinative: *tat-puruṣa* 'that - man'= 'the man of that one', 'his man'), and *bahuvrīhi* (possessive, or exocentric: *bahu-vrīhi* 'much - rice' = 'having much rice').

Syntax

Sanskrit has free word order, but in unmarked statements displays a tendency toward SOV order; unaccented pronouns and other clitics are consistently placed after the first accented word in the sentence in which they occur (Wackernagel's position), modifying adjectives, genitives, and restrictive relative clauses precede their head, and adpositions mostly follow the noun. OV features are all present in Vedic, too, but Classical Sanskrit is more consistent with respect to them.

Subordination includes relative clauses and a variety of adverbial clauses, with conjunctions derived from the stem of the relative pronoun. Besides, participles and gerunds also contribute to the building of hypotaxis. Participles (especially the past passive participle in *-ta*) and gerunds are often used without the verb 'be', as if they were finite verb forms.

Sanskrit displays a peculiar preference for passive, rather than active construction. The would-be-subject of the active is thus expressed through an agent phrase in the instrumental: *tenedam uktam*, 'he said this', lit.: 'this (was) said by him' (*tena*, 'by him' + *idam*, 'this', *uktam*, past passive participle of *vac-*, 'say'). Impersonal passives with instrumental agent phrases are also found in intransitive verbs; the past participle, however, occurs more frequently with an agent–subject phrase in the nominative, in which case it has active meaning (so *saḥ grāmam gataḥ*, 'he went to town', *sah*, 'he', *grāmam*, 'to-town', *gataḥ*, past participle of *gam-*, 'go', rather than *tena grāmam gatam*, lit.: 'by him (*tena*, instr.) (it was) gone to town').

Sanskrit and Indo-European

The importance of Sanskrit for Indo-European linguistics cannot be overstated: it was only after European scholars learned about it that they realized that their own European languages presented striking similarities among each other, and set out to reconstruct a proto-language. Actually, for some time Sanskrit was even thought to be the proto-language: but also later, after it lost this status, Sanskrit was regarded as the most conservative Indo-European language, especially with respect to its numerous inflectional categories, fully developed in rich and complex paradigms.

The discovery of Anatolian, at the beginning of the twentieth century, gave a serious stroke to the primacy of Sanskrit. Some scholars started to reconstruct a Proto-Indo-European with a smaller case system, with less verbal moods, and in which some basic Greek–Sanskrit isoglosses (augmented past tenses, aorist, dual number, etc.) played a less relevant role. Another challenge to the possible archaic character of Sanskrit has come from the so-called 'ejective model', according to which the stops of Proto-Indo-European include voiceless/glottalized/aspirate, rather than voiceless/voiced/voiced aspirate of the traditional reconstruction (see Hopper 1973). A full discussion of new reconstruction trends can be found in Gamkrelidze and Ivanov (1995).

References

Allen, Sidney. 1962. Sandhi. The Hague: Mouton.

Bloch, J. 1965. Indo-Aryan from the Vedas to modern times. Paris: Adrien-Maisonneuve.

Burrow, T. 1965. The Sanskrit language. London: Faber & Faber.

Cardona, George. 1990. Sanskrit. The world major languages, ed. by B. Comrie. Oxford: Oxford University Press.

Gamkrelidze, Th. V., and V. V. Ivanov. 1995. Indo-European and the Indo-Europeans. Berlin, New York: Mouton de Guyter.

Gonda, Jan. 1951. Remarks on the Sanskrit passive. Leiden: Brill.

Hoffman, Karl. 1967. Der Iniunktiv im Veda. Heidelberg: Winter.

Hopper, Paul. 1973. Glottalized and murmured occlusives in Indo-European. Glossa 7(2). 141–66.

McDonell, A.A. 1916. A Vedic grammar for students. Oxford: Oxford University Press.

McDonell, Arthur A. 1926. A Sanskrit grammar for students. Oxford: Oxford University Press.

Speijer, J. S. 1980. Sanskrit syntax. Delhi: Motilal Banarsidass (reprinted from 1886 edition).

SILVIA LURAGHI

Sapir, Edward

As a senior in the undergraduate program in German language and literature at Columbia University in 1903–1904, Edward Sapir enrolled in a graduate seminar in American languages taught by the distinguished anthropologist Franz Boas. Sapir had already studied Latin, Greek, French, Spanish, and his major subject German, and now he was exposed to the diversity of the native languages of North America. The effect was not immediate, for Sapir continued on to the master of arts degree, still concentrating on Germanic philology and literature. But his 1905 master's thesis, on traditional German linguistics, did include data from several American languages to demonstrate the grammatical complexity of all languages, those without writing systems as well as those with long established literary traditions. Soon Sapir was engaged in fieldwork, traveling to the West Coast of the United States to study Wishram in Washington in the summer of 1905, and Takelma in Oregon in 1906. Although his Germanic studies had been largely philological and historical, his grammar of Takelma was entirely a synchronic analysis, describing the language without recourse to historical information, a characteristic of most of the linguistic work done under the direction of his mentor, Franz Boas.

Sapir published his master's thesis, in the journal *Modern Philology,* but virtually all of his other early publications were on native American languages. His doctorate was awarded in 1909, with the Takelma grammar serving as his dissertation, and by the end of that year he had published nine articles, reviews, and monographs on the Kwakiutl, Upper Chinook, Yana, Yuchi, and Wishram languages, as well as Takelma. His reputation as a specialist in native American languages was established early, and continued to grow throughout his lifetime, reaching nearly mythic proportions as some admirers attributed to him extensive knowledge, and even fluent command, of languages which he had investigated for just a few days in consultation with a single speaker.

After receiving the doctorate, he took up the post as chief of the Division of Anthropology in the Geological Survey of the Canadian National Museum. This was both a hindrance and an asset to his career as a linguist. On the one hand, the position isolated him from his colleagues in the United States and at times directed his energies toward administration. But the advantages were many, not least of which was control of a research staff and a budget, always too small, with the authority to determine and direct fieldwork. It was Sapir's intent to create a systematic survey of the languages and cultures of all of Canada. Since he was nearly always the only staff member with formal training in linguistics, it was he who did most of the linguistic work, concentrating his attention on the languages of Western Canada. He made many field trips in the years 1910–1915, and thereafter worked frequently with speakers at his office in Ottawa.

The use of trained native assistants for the recording of text was a common practice with Boas and his students, but Sapir was exceptional in recognizing that native speakers' insights were relevant to the development of linguistic theory, especially in the area of phonology. In the early years of Sapir's career, linguists were struggling with the new concept of the phoneme, which Sapir later came to define as

'a functionally significant unit in the rigidly defined pattern or configuration of sounds peculiar to a language . . ., as distinct from . . .the 'sound' or 'phonetic element' as such (an objectively definable entity in the articulated and perceived totality of speech)'. (Sapir, 'The psychological reality of phonemes' 1985[1933]: 46)

One of Sapir's most important and long-lasting contributions to the development of the phoneme concept was his view that the phonemic structure of a language was part of the unconscious knowledge that speakers of a language possessed, part of their 'phonological intuitions.' Drawing on his years of work with native interpreters of American Indian languages, he argued that ordinary speakers hear phonemes, not phonetic details, and he offered evidence in the way that interpreters would transcribe their language, always recording distinctive contrasts in sounds while not noting phonetic differences that were irrelevant to their phonological system. The paper in which these concepts were most clearly expressed was written in French and published in Paris, France in 1933. At that time, it was little noted by American linguists, but when it appeared in an English version in 1949, ten years after Sapir's death, it sparked something of a revival of interest in Sapir's work. His notions of the psychological reality of the phoneme, of native speaker intuition, and of unconscious linguistic knowledge were all to find their way into the development of generative phonology in the late 1950s and 1960s, as did his description of phonological systems as 'sound patterns', explicated in his widely read 1925 article 'Sound patterns in language' published in the first issue of *Language*, journal of the Linguistic Society of America.

Also growing out of Sapir's work in Canada was his genetic classification system for American Indian languages. Drawing on his background in Germanic historical linguistics, Sapir began to apply the techniques of the Indo-European comparative method to the original languages of America. Beginning with the 1891 classification system developed by John Wesley Powell and the Bureau of American Ethnology, Sapir took Powell's 55 'stocks' (or language families) and eventually reduced them to six groups: Eskimo-Aleut, Algonquian-Ritwan, Na-Dene, Penutian, Hokan-Siouan, and Aztec-Tanoan. There was a good deal of controversy at the time over this plan, in part because Sapir provided little evidence to support it. Sapir himself considered the scheme to be probable but not at all proven, given that 'scientific comparative work on these difficult languages is still in its infancy' ('Central and North American languages', 1985[1929]: 169), but his system remains the fundamental reference point for those working on the genetic classification of American Indian languages.

Sapir's most widely read work is the small book *Language*, published in 1921 while he was still in Canada. Aimed at both students of linguistics and the public, the volume predated his work on phonology. Its chapters focus on words and grammar, with a particular emphasis on what is today called linguistic typology, the classification of languages and of features of languages into types based on the structural properties of their units (as opposed to genetic classification based on shared history). Sapir's classification system drew heavily on his background in Germanic linguistics and can be viewed as the twentieth-century apex of nineteenth-century European typology. The system provides categories for the morphology (the processes of word formation) of languages along two dimensions: the extent to which words employ affixes and the ease with which the root and affixes of a word can be segmented. For example, an isolating language would be one in which most words consist solely of roots, with little affixing. Sapir gives Chinese as an example. A language with many affixes intertwined and difficult to segment would be labeled fusional; Sapir cites Salinan, a language of southwest California. This part of Sapir's typology, based as it was on the well–known categories of earlier linguists, has been modified somewhat in terminology, but it remains useful in the classification of morphological types. Sapir, however, also attempted to overlay this system with another, a conceptual classification that distinguished concrete concepts and relational concepts, with further bifurcation into simple and complex. This aspect of his typological classification has not been pursued, but Sapir himself recognized the preliminary nature of his system and the need to explore a greater number of more diverse languages than he was able to consider at the time. This is exactly what modern typologists have done.

Sapir's interest in conceptual categories was manifested in another area of research in which he engaged in the later 1920s and early 1930s—the issue of an international auxiliary language. Beginning in 1925, at about the time he accepted a position at the University of Chicago, Sapir became involved with the International Auxiliary Language Association, Inc. (IALA), headquartered in New York City. Working in cooperation with the co-founder and honorary secretary of the association, Alice Vanderbilt Morris, Sapir wrote a number of scholarly and popular pieces supporting research on and development of an international auxiliary language. He was not alone among linguists in this interest, which was shared most enthusiastically by the Danish linguist Otto Jespersen, a number of other European linguists, as well as by many of Sapir's American colleagues. Sapir served briefly as IALA's linguistic research director and under IALA sponsorship produced three studies of conceptual structure, which comprise Sapir's total output on semantics. He made a number of programmatic statements but did no further work in this area and has had no lasting influence. Linguists' interest in an international auxiliary language dissipated during the 1930s, a period of professionalization within the discipline and of increasing American isolationism on the national scene.

In 1931, Sapir moved from the University of Chicago to Yale University. He never again conducted fieldwork, and he largely withdrew from linguistic study to concentrate on culture and personality. But he did supervise dissertations and postdoctoral work on American languages by a number of Yale students who went on to distinguished careers in linguistics, including Mary R. Haas, Stanley Newman, Morris Swadesh, and Charles F. Voegelin. At Yale, Sapir also came to know Benjamin Lee Whorf, a largely self-educated man interested primarily in philosophical issues and the study of meaning, who associated himself with Sapir and his circle of graduate students. Whorf's linguistic relativity hypothesis (that grammatical structure of a language influences its users' perceptions of the world) became well known beyond the field of linguistics. It is sometimes referred to, incorrectly, as 'the Sapir–Whorf hypothesis', but it was developed fully only after Sapir's death in 1939. Most contemporary linguists found it too speculative, and the hypothesis never became part of mainstream linguistics.

Edward Sapir was a key figure of American linguistics in the first half of the twentieth century, equaled in stature for that period only by his contemporary Leonard Bloomfield. Sapir's students protected and promoted his work and ideas and he remains 'a symbol of scope and insight in the study of language' (Dell Hymes 1985: 598).

Biography

Edward sapir was born on January 26, 1884 in Lauenberg, Germany, now part of Poland, the son of Lithuanian parents who soon moved to England and then on to the United States. He spent his childhood in Richmond, Virginia up to the age of ten, and then in New York City. He began the study of German, French, Spanish, Latin, and Greek at Peter Stuyvesant High School, matriculating with a German major at Columbia University in 1901. He received his B.A. in 1904; M.A. (1905) in Germanic philology and literature; and Ph.D. (1909) in anthropology, specializing in linguistics with a dissertation on the Takelma language of Oregon supervised by Franz Boas. He was Research Associate, Department of Anthropology, University of California, Berkeley, 1907–1908, and Fellow and then instructor at the University of Pennsylvania, 1908–1910. He was Chief, Division of Anthropology, Geological Survey of the Canadian National Museum, Ottawa, 1910–1925. He went to the University of Chicago in 1925, and was promoted to Professor of Anthropology and General Linguistics in 1927. Concurrently, he was also Director of Linguistic Research, International Auxiliary Language Association, New York City, from October 1930 to July 1, 1931. He went to Yale University as Sterling Professor of Anthropology and Linguistics in 1931. He was President, Linguistic Society of America (LSA) in 1933; he also taught Introduction to Linguistic Science and Field Methods at the 1937 summer LSA Linguistic Institute at University of Michigan. Sapir died in New Haven, Connecticut on February 4, 1939.

References

Anderson, Stephen R. 1985. Phonology in the twentieth century: theories of rules and theories of representations. Chicago and London: University of Chicago Press.

Campbell, Lyle. 1997. American Indian languages: the historical linguistics of native America. New York and Oxford: Oxford University Press.

Cowan, William, Michael K. Foster, and Konrad Koerner (eds.) 1986. New perspectives in language, culture, and personality: proceedings of the Edward Sapir centenary conference, Ottawa, 1–3 October 1984. Amsterdam and Philadelphia: John Benjamins.

Darnell, Regna. 1990. Edward Sapir: linguist, anthropologist, humanist. Berkeley and Los Angeles: University of California Press.

Falk, Julia S. 1995. Words without grammar: linguists and the international auxiliary language movement in the United States. Language & Communication 15(3). 241–59.

Hymes, Dell. 1985. Epilogue. Selected writings of Edward Sapir in language, culture and personality, ed. by David G. Mandelbaum, paperback edition, 598–600. Berkeley and Los Angeles: University of California Press.

Koerner, Konrad (ed.) 1984. Edward Sapir: appraisals of his life and work. Amsterdam and Philadelphia: John Benjamins.

Lee, Penny. 1996. The Whorf theory complex: a critical reconstruction. Amsterdam and Philadelphia: John Benjamins.

Sapir, Edward. 1989. The collected works of Edward Sapir, 16 vols. Berlin and New York: Mouton de Gruyter.

Sapir, Edward. 1921. Language: an introduction to the study of speech. New York: Harcourt, Brace.

Whaley, Lindsay J. 1997. Introduction to typology: the unity and diversity of language. Thousand Oaks, CA: Sage Publications.

All the following works are reprinted in Selected writings of Edward Sapir in language, culture and personality, ed. by David G. Mandelbaum. Berkeley and Los Angeles: University of California Press, 1949. Page references in the text are to the paperback edition of 1985.

Sapir, Edward. 1925. Sound patterns in language. Language 1. 37–51.

———. 1929. Central and North American languages. Encyclopaedia Britannica, 14th edition, Vol. 5, 138–41. London and New York: Encyclopaedia Britannica Co.

———. 1929. The status of linguistics as a science. Language 5. 207–14.

———. 1931. Function of an international auxiliary language. Psyche 11. 4–15.

———. 1933. Language. Encyclopaedia of the social sciences, Vol. 9, 155–69. New York: Macmillan.

———. 1933. La Réalité Psychologique des Phonèmes. Journal de Psychologie Normale et Pathologique 30. 247–65.

———. 1944. Grading: a study in semantics. Philosophy of Science 11. 93–116.

JULIA S. FALK

See also **Boas, Franz; Sapir–Whorf Hypothesis; Whorf, Benjamin Lee**

Sapir–Whorf Hypothesis

The idea that one's native language 'colors' his *Weltanschauung* has been around at least since the days of Johann Gottfried Herder (1744–1803) and Wilhelm von Humboldt (1762–1835). Charles Sanders Peirce (1839–1914) postulates that man's symbolic universe could only make sense via language. Ferdinand de Saussure (1857–1913) *(Cours de linguistique générale* [1916:155]) states that: 'No ideas are established in advance, and nothing is distinct, before the introduction of linguistic structure.' However, the principle of linguistic relativity has largely become associated with Benjamin Lee Whorf (1897–1941), who along with Edward Sapir (1884–1939), his linguistic mentor at Yale University, used linguistics to advocate their position that language influences the way in which a speech community perceives and conceives of its reality.

Part of the groundwork for the Sapir–Whorf hypothesis was laid by Whorf's work as a fire insurance investigator. During his career, he had the opportunity to analyze many reports as to why fires broke out. He found that workers would use extreme caution when around 'full' drums of gasoline. Just as one might expect, workers were careful not to smoke around 'full' drums. Yet, these same workers, when around 'empty' drums of gasoline, would sometimes toss lit cigarettes nearby. This caused a violent explosion because an empty drum still contained volatile gasoline vapor. Thus, an 'empty' drum was really much more of a threat than a 'full' one. Using these data, Whorf concluded that the meanings of certain words had an effect on a person's behavior.

It was the extensive research of both Sapir and Whorf into the grammatical systems of American Indian languages, however, that proved to have the greatest impact on the hypothesis. By predicating their insights into the interrelationships of language and culture on what they had learned from the structures of these languages, the basic idea of language shaping the perceptions of its speakers and providing for them a vehicle so that their experiences and emotions could be placed into significant cognitive categories was given its scientific underpinnings. Generally, Sapir is credited as giving the problem of establishing the link between language and culture its initial formulation, while Whorf is honored as the one who took this idea and developed it further to include grammar in addition to lexis, thereby making it a bona fide hypothesis. Hence, the resultant supposition is commonly given

the designation the 'Whorfian hypothesis'. Some, pointing to Sapir's preeminent stature as a linguist, prefer the appellation the 'Sapir-Whorf hypothesis'. When viewed in terms of output, however, one could counter that a more appropriate label would be the 'Whorf–Sapir hypothesis'.

A rather interesting development in this debate over giving credit where credit is due has been the attempt to disassociate Sapir from the hypothesis entirely. Desirous of preventing the image of the great maestro Sapir from being tarnished by the taint of controversy, some, most notably Alfred L. Kroeber, have claimed that Sapir's views were not really that (pro)Whorfian. This viewpoint is not borne out by an examination of Sapir's own writings. For example, as one can see in a (1929) passage (from the journal *Language*, p. 209), he equated language and thinking in terms of the speech community's overall culture:

> Language is a guide to 'social reality'…it powerfully conditions all our thinking about social problems and processes. Human beings do not live in the objective world alone, nor alone in the world of social activity as ordinarily understood, but are very much at the mercy of the particular language which has become the medium of expression for their society.

There are actually two different versions of the hypothesis. This is understandable when one considers that Whorf did all of his writing from 1925 to 1941, and his ideas were continuously evolving. The strong version of the hypothesis, which is called linguistic determinism, holds that language determines thinking. This position is most difficult to defend primarily because translation between one language and another is possible, and 'thinking' can take place without language; e.g. an artist or sculptor can and often does think with his fingers.

Mirroring Sapir's thoughts, Whorf notes in his (1940) 'Science and linguistics':

> We dissect nature along lines laid down by our native languages…We cut nature up, organize it into concepts, and ascribe significances as we do, largely because we are parties to an agreement to organize it in this way — an agreement that holds throughout our speech community and is codified in the patterns of our language. (Carroll 1956:213).

The milder version of the hypothesis is labeled linguistic relativity, coined by Whorf, since he always tried

to qualify his assertions. This claims that the native language *influences* one's thoughts and perceptions.

Linguistic relativity can be illustrated using one of Whorf's favorite sources of data, (Uto-Aztecan) Hopi. With the exception of birds, there is, according to Whorf, only one word in Hopi for everything that flies. One also notes that Bedouin Arabic dialects have more than 6,000 different lexemes for various types of camels and camel paraphernalia. Perhaps the most incontrovertible piece of evidence in its favor comes from the realm of numbers. Some languages (e.g. Hottentot [Nama]) only have words for the numerals 'one' and 'two' and a word roughly translatable as 'many' for 'three or more' (and a few languages have no numbers at all). Thus, such concepts as googolplex are beyond everyday verbal expression in these languages. Mathematics (as we know it) is thus not possible, i.e. people are indeed 'at the mercy of' their native tongues.

In 'An American Indian Model of the Universe' (*International Journal of American Linguistics* [1950]), Whorf argues that since there is neither an explicit nor an implicit reference to time in Hopi and thus no tenses for its verbs, time and space cannot be the same concepts for Hopis as they are for English speakers. The Hopi *Weltanschauung* is different from that of an SAE (Standard Average European, a Whorfian term) speaker. Many of his other essays present data for his basic contention that Hopi metaphysics, which underlies its cognition, is different from that of an English speaker; i.e. the Hopis calibrate the world differently because their language defines experience differently for them. As more information has surfaced on Hopi, some of Whorf's specific grammatical points have not held up well. This explains why most linguists today believe the Sapir–Whorf hypothesis to be invalid. However, the truth of the matter is that while no one has proved it wrong, neither has anyone proved it right.

Whichever end of the continuum one considers in the relationship between language and culture, it is important to realize the interpenetration of the two. In areas like bilingualism, is it really possible to learn a foreign language without also simultaneously learning the *Weltanschauung* of its speakers? Not only is language part of culture, it allows for its acquisition.

Sapir maintains that 'the worlds in which different societies live are distinct worlds, not merely the same world with different labels attached' (Mandelbaum 1949:162). Whorf took that one step further, asserting that grammar not only allows for the voicing of ideas but it also 'is the shaper of ideas' (Carroll 1956:221).

References

Carroll, John B. (ed.) 1956. Language, thought, and reality: selected writings of Benjamin Lee Whorf. Cambridge, MA: MIT Press.

Darnell, Regna. 1990. Edward Sapir: linguist, anthropologist, humanist. Berkeley: University of California Press.

Gumperz, John J., and Stephen C. Levinson (eds.) 1997. Rethinking linguistic relativity. Cambridge: Cambridge University Press.

Hoijer, Harry (ed.) 1954. Language in culture. Chicago, IL: University of Chicago Press.

Koerner, Konrad (ed.) 1984. Edward Sapir: appraisals of his life and work. Amsterdam and Philadelphia: John Benjamins.

Lee, Penny. 1996. The Whorf theory complex. Amsterdam and Philadelphia: John Benjamins.

Mandelbaum, David G. (ed.) 1949. Selected writings of Edward Sapir in language, culture, and personality. Berkeley: University of California Press.

ALAN S. KAYE

See also **Humboldt, Wilhelm von; Peirce, Charles; Sapir, Edward; Saussure, Ferdinand de; Whorf, Benjamin Lee**

Saramaccan

Saramaccan (or Saamáka) is the name of the creole language spoken by the Saramaka people, who live along the Suriname River in central Suriname (in the northeast of South America). The names 'Saramaka' and 'Saamáka' derive from the fact that the first settlements of these people were located along the Saramacca River, also in central Suriname. The c. 50,000 Saramaccans living today are the descendants of African slaves who escaped from the plantations to create their own communities in the Suriname rain forest in the seventeenth and eighteenth centuries. These runaway slaves and their descendants are often referred to as 'Maroons', a word derived from Spanish *cimarron*, meaning 'stray animal'. The creole languages spoken by Maroon communities may be referred to as 'maroon creoles', to distinguish them from (former) 'plantation creoles', such as Sranan. There are two Maroon creoles (or rather: groups of Maroon creoles) in Suriname: Saramaccan (including three different dialects) and Ndyuka (used here as a cover term for four different dialects). Maroon

creoles outside of Suriname that are still spoken today include Palenquero (Colombia) and Angolar (São Tomé, an island off the coast of West Central Africa).

The main reason for distinguishing Maroon creoles as a separate category is the fact that, due to their relative isolation from outside influence, they are assumed to be more 'radical' than (former) plantation creoles, which have remained in contact with the language from which they have derived their lexicon (the 'lexifier language'). The term 'radicality' refers to the typological distance between a creole and its lexifier language. Although until now very little comparative research regarding the degree of radicality of different creoles has been done, it seems clear that the typological distance between, say, Saramaccan and its (main) lexifier, English, is larger than that between, say, Cape Verdean Creole and Portuguese. Therefore, Maroon creoles are assumed to be closer to creoles at the time of their formation (say, 300 years ago) than (former) plantation creoles. Among the Maroon creoles, Saramaccan has acquired a special status, especially in research associated with Bioprogram Theory (see e.g. Bickerton 1984), to the extent that it has come to be considered as the most radical creole. According to this view, Saramaccan provides a rare opportunity to get a closer look at the process of creolization.

While there are still important gaps regarding the history of the Saramaka and their language, the following seems to be clear (Smith 2002). The origins of the Saramaka people and their language go back to escapes from plantations that took place before 1700, at a time when the formation of a plantation creole must have been well on its way. Assuming that most of the runaway slaves had been on the plantations for some time before making their escape, they took at least some knowledge of this evolving plantation creole with them. This explains the structural similarities between Saramaccan and Sranan, both of which are descendants of the earlier plantation creole.

In spite of these similarities, however, there are also a number of differences between the two languages. One of these is the proportion of Portuguese-derived words, which is much larger in Saramaccan than it is in Sranan. In the former, one third of the basic vocabulary (basic words such as 'water', 'go', 'big') is derived from Portuguese, while this is much smaller in Sranan. The remainder of the basic vocabulary is largely derived from English (Suriname was an English colony before it became Dutch), while there are also a few items from West African languages. The presence of these Portuguese-based words is explained by the fact that, as has been established by historical research (e.g. Price 1976, 1983), many of the first Saramaka came from plantations owned by Sephardic Jews, who were Portuguese- (and to some extent Spanish-) speaking. Although the presence of many Portuguese-derived words could lead one to view Saramaccan as a creole with two lexifier languages—English and Portuguese—the fact that most function words are from English suggests that the Portuguese element was added later. For this reason, Saramaccan is generally categorized as an English-lexicon creole, albeit one with a strong Portuguese element.

A second difference between Saramaccan and Sranan is related to the fact that the former has a higher percentage of words derived from African languages. This is probably due to the fact that there was much less contact with other languages in the case of Saramaccan than in the case of Sranan. As to the origins of these words, three African languages (or rather: language clusters) have been especially important: Kikongo (a Bantu language), Gbe, and Akan (both of them Kwa languages; both Bantu and Kwa belong to the family of Niger-Congo languages). Although nothing is known about the specific African origins of the individual runaway slaves who formed the 'founder population' of the Saramaka people, we do have reliable information about the origins of the African slaves in general, who were brought to Suriname in the 1675–1700 period (Arends 1995a:243). In this period, roughly speaking, half of all Suriname slaves came from an area where Bantu languages such as Kikongo, were spoken, while the other half came from an area where Kwa languages, such as Gbe and Akan, were spoken. The connection between the ethnolinguistic origins of the Suriname slaves and the traces that were left in the Suriname creoles by their native languages receives further support from the fact that Saramaccan exhibits some rather marked phonological features, such as lexical tone and nasal and complex stops, which are characteristic of these three languages (Bruyn 2002).

Like its sister language Sranan, Saramaccan is well documented in its early stages. In the case of Saramaccan, however, the early documentation is limited to a very short period, roughly 1780–1820. This has to do with the fact that the Moravian Brethren, to whom we owe these early writings, stopped their missionary activities among the Saramaka around 1820. These early documents, which together number well over 2,000 manuscript pages, consist mainly of religious texts, such as Bible translations, although some linguistic descriptive works, such as dictionaries, are included as well (see Arends (1995b) for further information). Unfortunately, however, up to now only a few of these documents have been made available for linguistic research (Arends and Perl 1995).

As to the major structural features of Saramaccan, many of these are also found in Sranan. Here, we will only mention and illustrate some of the features in which the two languages differ. (This section draws heavily on Bruyn 2002; see also Bakker et al. (1995.)

Lexicon. Some examples of Portuguese-derived basic vocabulary items are *búka* (*<boca*) 'mouth' and *dá* (*<dar*) 'give'. In both cases, the equivalent word in Sranan is derived from English: *mofo* (*<mouth*) 'mouth' and *gi* (*<give*) 'give'. Some examples of African-derived words are *katangá* 'cramp' from Kikongo *nkatangá*, *azɛ* 'magic' from Gbe *àze*, and *gongosá* 'gossip' from Twi *ŋkɔŋkɔnsá*.

Phonology. For a language to have lexical tone, as does Saramaccan, means that some words differ only in their tonal pattern, for example in having a low tone in a syllable where the other word has a high tone, while being completely identical in every other respect. An example is the pair *ná~nà*, with the first having a high tone where the second has a low. Despite this (seemingly) small difference, the meaning of the two words is completely opposite: *ná* means 'be', while *nà* means 'be not'. Examples of nasal stops are /mb/ and /nd/; examples of complex stops are /kp/ and /gb/. Note that /mb/, etc., refer to phonemes, not combinations of phonemes. In other words, a word like *mbéti* 'meat, animal' consists of four phonemes, not five.

Morphology. Apart from other functions, reduplication is used in Saramaccan to derive adjectives from verbs, e.g. *nákináki* 'beaten', derived from *náki* 'beat'. These reduplicated forms are used both attributively, as in *dí nákináki miíi* 'the beaten child', and predicatively, as in *dí miíi dɛ nákináki* 'the child has been beaten (is in a beaten state)'. Saramaccan also differs from Sranan in that the agentive suffix *-ma* (*-man* in Sranan) may follow an entire verb phrase, which may itself even contain a subordinate clause. This may result in quite complex agentive nouns, such as *seti-u-kanda-ma* 'precentor' (lit. 'start-to-sing-suffix').

Syntax. The syntactic differences between Saramaccan and Sranan have not been investigated in sufficient detail to allow them to be discussed here. One feature that has drawn considerable attention, however, is the serial verb construction, which is also found in Sranan. For an elaborate discussion of serial verbs in Saramaccan, the reader is referred to Veenstra (1996).

Verbal arts. In the literature on the 'verbal arts' (story telling, song, etc.), Saramaccan is often discussed together with Sranan (Herskovits and Herskovits 1936; Voorhoeve and Lichtveld 1975). A work that is entirely devoted to storytelling in Saramaccan is *Two evenings in Saramaka* by Richard and Sally Price (1991).

References

Arends, Jacques. 1995a. Demographic factors in the formation of Sranan. The early stages of creolization, ed. by Jacques Arends, 233–85. Amsterdam and Philadelphia, PA: Benjamins.

———. 1995b. Early Saramaccan texts in the Moravian archives in Paramaribo, Herrnhut and Utrecht. The Carrier Pidgin 23(3). 1–6.

Arends, Jacques and Matthias Perl. 1995. Early Suriname creole texts: a collection of 18th-century Sranan and Saramaccan documents. Frankfurt: Vervuert.

Bakker, Peter, Norval Smith, and Tonjes Veenstra. 1995. Saramaccan. in Pidgins and creoles: an introduction, ed. by Jacques Arends, Pieter Muysken, and Norval Smith, 165–78. Amsterdam and Philadelphia, PA: Benjamins.

Bickerton, Derek. 1984. The language bioprogram hypothesis. The Behavioral and Brain Sciences 7. 173–88.

Bruyn, Adrienne. 2002. The structure of the Surinamese creoles, Carlin and Arends. 153–82.

Carlin, Eithne, and Jacques Arends (eds.) 2002. Atlas of the languages of Suriname. Leiden: KITLV Press.

Herskovits, Melville, and Jean Herskovits. 1936. Suriname folk-lore. New York: Columbia University Press.

Price, Richard. 1976. The Guiana Maroons: a historical and bibliographical introduction. Baltimore: The Johns Hopkins University Press.

———. 1983. First-time: the historical vision of an Afro-American people. Baltimore: The Johns Hopkins University Press.

Price, Richard, and Sally Price. 1991. Two evenings in Saramaka. Chicago: Chicago University Press.

Smith, Norval. 2002. The history of the Suriname creoles: origin and differentiation, Carlin and Arends. 131–51.

Veenstra, Tonjes. 1996. Serial verbs in Saramaccan: predication and creole genesis. Unpublished Dissertation. University of Amsterdam.

Voorhoeve, Jan, and Ursy Lichtveld (eds.) 1975. Creole drum: an anthology of Creole literature in Suriname. New Haven and London: Yale University Press.

JACQUES ARENDS

Saussure, Ferdinand de

Ferdinand de Saussure is known as one of the founders of contemporary linguistics and as the father of European Structuralism. Throughout his life he retained a major interest in historical linguistics, his student dissertation on the *a* vowels of Indo-European (1879) containing the solution to a problem that was to

be empirically confirmed many years later. Certain concepts for which he was to become known were present in his early works, but would be more fully developed in his Geneva lectures, notes from which were published posthumously in 1916 as the *Cours de linguistique générale (Course in general linguistics (CLG))*. In the first three sections of the *Cours*, Saussure attempts to come to grips with the kind of phenomenon that we are studying when we analyze language. In the face of the many possible approaches (historical, sociological, etc.) to the heterogeneous phenomenon that is 'langage' (translated as 'language' or as 'language faculty'), Saussure tries to focus on its essence through a series of binary divisions. First, it is necessary to distinguish the underlying system *('langue')* from its individual representations *('parole')*. Second, Saussure differentiates between the diachronic (evolution of a language) and the synchronic (a snapshot of a language at a particular time), the latter being the proper object of study in general linguistics. 'Langue' is a system in which the whole is more than the sum of the parts (Saussure himself did not use the term structure). The system is made up of units, each of which is defined by its difference from other units, whether words or phonemes (it should be noted Saussure generally used 'phonology' where today we would use 'phonetics' and vice versa). Each unit functions as a linguistic 'sign' composed of two inseparable faces: the *'signifiant'* (signifier or acoustic image) and the *'signifié'* (signified or mental representation of external reality). It is a characteristic feature of sign systems, of which language is the principal one, that the link between the sign and the reality to which it refers is arbitrary. Linguistic signs are linked together either paradigmatically (one is selected from a series, e.g. grammatical cases) and syntagmatically (in a chain within the utterance).

In talking about 'langue', Saussure refers to a mental reality, and to a collective and therefore social phenomenon. That he considers linguistics proper to be a discipline in its own right and as such separate from psychology and sociology does not, however, mean that he denies the links between language and history or society. He has been criticized for neglecting 'parole', syntactic analysis, and units beyond the sentence. It appears from the *CLG* that he intended a 'linguistique de la parole' to follow the 'linguistique de la langue', and that he considered syntax to fall within the remit of 'parole'. The discipline—then still to be established—of which Saussure considered linguistics to be the most important part is semiology (otherwise known as semiotics).

There has been much debate about influences on Saussure's theories, but it is generally agreed that he provides key insights that interlock to form the basis of a new and coherent theory. There is no doubt about the influence of the *CLG* on European Structuralist Linguistics and on Structuralist thought more generally. Firstly, through his lectures in Paris, Saussure exerted a major influence on future generations of French linguists, through Meillet for historical linguistics and Passy and Grammond for phonetics, to Benveniste and Guillaume; in Switzerland, the Geneva School continued his work, with Bally working on a 'linguistique de la parole'. Through his writing and through the presence of Karcevski, he influenced the Prague Circle including Jakobsen, as he did the Copenhagen Circle. In the English-speaking world, his influence existed, but in a weaker form: Bloomfield published a positive review of the *Cours* in 1924, and it is an important point of reference in Chomsky's *Cartesian Linguistics* (with langue/parole being in some ways a precursor of competence/performance). For the Structuralists of the mid-twentieth century in other disciplines, as for the postmodernists, the *CLG* is the fundamental text. Saussure's work has been continuously promoted by the *Cahiers Ferdinand de Saussure* and there has been a regeneration of interest with the ongoing publication of new manuscript material and the creation of an Institut Ferdinand de Saussure. When reading the *Cours*, it is important to remember that Saussure himself did not write it, and also that the English translations of some of his terms are misleading. Thus, the ideas of a major twentieth-century thinker come down to us in a tantalisingly uncertain form.

Biography

Saussure was born in Geneva in 1857 into a family of distinguished scientists and scholars. In 1876, he went to study in Leipzig, where he encountered the 'neogrammarian' movement. In 1877, he presented his first paper to the Société de Linguistique de Paris, and in 1879 published a dissertation entitled *Mémoire sur le système primitif des voyelles des langues indo-européennes,* which immediately brought him renown. The following year, his doctorate on the genitive in Sanskrit was awarded *summa cum laude* and published in 1881. After a field trip to Lithuania in 1880, Saussure went to Paris to study at the Ecole Pratique des Hautes Etudes under Bréal and from the following year began to teach there. In 1891, Saussure returned to Geneva to take up a Chair in Indo-European linguistics, and later in General Linguistics. In three series of lectures (1907, 1908–1909, and 1910–1911), he covered historical linguistics and started to lay the foundations for the science of linguistics. In contrast to the brillance of his early years, in later life Saussure published relatively little, apparently feeling convinced of the need to forge the basic concepts of linguistic analysis but daunted by this

prospect. He developed related interests, writing about the possible widespread presence of anagrams in Latin poetry and about the German epic *Niebelungenlied*. In addition to this crisis of confidence, Saussure suffered from ill health, and died in 1913.

References

Arrivé, M. 2000. Saussure. Paris: Les Belles Lettres.

Bouquet, Simon. 1997. Introduction à la lecture de Saussure. Paris: Payot.

Fehr, Johannes. 2000. Saussure entre linguistique et sémiologie. Paris: Presses Universitaires de France.

Gadet, Françoise. 1989. Saussure and contemporary culture. London: Hutchinson.

Gordon, W. Terence. 1996. Saussure for beginners. New York: Writers and Readers.

Harris, Roy. 1987. Reading Saussure. London: Duckworth.

———. 2001. Saussure and his interpreters. Edinburgh: Edinburgh University Press.

Holdcroft, David. 1991. Saussure. Signs, system and arbitrariness. Cambridge/New York: Cambridge University Press.

Koerner, E.F.K. 1999. Linguistic historiography projects and prospects, Chapters 6, 8, 10. Amsterdam/Philadelphia: John Benjamins.

Normand, Claudine. 2000. Saussure. Paris: Les Belles Lettres.

Sanders, Carol. 2000. Saussure translated. Historiographia Linguistica XXVI(2/3). 345–58.

———. 2005. The Cambridge Companion to Saussre. Cambridge: Cambridge University Press.

———. 2005. Writing in general linguistics, by Ferdinand de Saussure, ed. by S. Bouquet and R. Engler, and transl. by C. Sanders and M. Pires. Oxford: Oxford University Press.

Saussure, Ferdinand de. 1916. Cours de linguistique générale, publié par C. Bally et A. Sechehaye avec la collaboration de A. Riedlinger. Lausanne/Paris: Payot.

———. 1966. Course in general linguistics, transl. by Wade Baskin, 2nd edition. New York and London: Harper & Row.

———. 1968. Cours de linguistique générale, édition critique par R.Engler, Vol. I. Wiesbaden: Otto Harrassowitz, II.

———. 1983. Course in general linguistics, transl. by Roy Harris. London: Duckworth.

———. 1993. Troisième Cours de linguistique générale (1910–1911) d'après les cahiers d'Émile Constantin, ed./transl. by E. Komatsu and R.Harris. Oxford: Pergamon.

———. 1995. Cours de linguistique générale, publié par Charles Bally et albert Séchehaye; avec la collaboration de Albert Riedlinger; édition critique préparée par Tullio de Mauro. Paris: Payot.

———. 1996. Premier Cours de linguistique générale (1907) d'après les cahiers d'Albert Riedlinger, ed./transl. by E. Komatsu and G.Wolf. Oxford: Pergamon.

———. 1997. Deuxième Cours de linguistique générale (1908–1909) d'après les cahiers d'Albert Riedlinger et Charles Patois, ed./transl. by E. Komatsu and G.Wolf. Oxford: Pergamon.

Thibault, Paul J. 1997. Re-reading Saussure. London/New York: Routledge.

———. 2001. Ferdinand de Saussure: critical assessments of leading linguists. 6 vol. London/New York: Routledge.

CAROL SANDERS

See also **Jakobson, Roman; Meillet, Antoine; Structuralism**

Scandinavia

Scandinavia usually refers to the mainland Scandinavian countries Denmark, Norway, and Sweden. Denmark includes the Faeroes and Greenland, which are autonomous regions of the Danish kingdom. Finland includes the Swedish-speaking Åland Isles in the Baltic Sea. Finnish and Sami—the other main indigenous languages in the region—neither belong to the Nordic (Swe. *nordiska språk*) nor the Scandinavian languages. The general term for Denmark, Finland, Iceland, Norway, and Sweden is the Nordic countries (Swe. *Norden* or *de nordiska länderna*; Fi. *Pohjoismaat*, lit. 'the North countries'). When discussing Scandinavia and its languages, it is for historical and political reasons necessary to include all Nordic countries and the Samis. The Samis, the aboriginal people of the North, are inhabitants of Finland, Norway, and Sweden (and Russia). They form a supranational Sami homeland, Sapme, with its own linguistic and cultural characteristics. Finnish is also spoken in Finland, Norway, and Sweden, and Swedish is spoken in Finland.

Despite many types of language contacts throughout history, the Nordic countries have all attained a proportion of 90% or more speakers, for their dominant national languages, at the cost of inherent cultural and linguistic pluralism. Most speakers of the indigenous minority and secondary official languages of the Nordic countries are bilingual, and many are dominant in their second language, the majority language of their country.

Sweden (Sverige)

Sweden has a parliamentary and democratic monarchy, a population of nine million, and no official language *de jure*. Until 1809 it included Finland, and from 1814 to 1905 it formed a union with Norway.

Swedish is the first language of about 8.3–8.5 million people, Finnish (Sweden Finnish) has about 230,000 speakers (no official language censuses exist); Meänkieli (closely related to Finnish) has about 50,000 speakers, mainly in the North-Swedish Tornedalen region; a Finnish-speaking population in the central forest regions (Värmland, Dalecarlia) became extinct in the late 1960s, after four centuries. An equal amount of Finland Swedish descendants (migrants from Finland), Norwegian, and Danish speakers live in Sweden: about 50,000 each. For Finland, until recently, Sweden has been a buffer zone for cultural influence from Southern and Western Europe. Similarly, Finland has functioned as a buffer zone for Swedish contacts—peaceful and hostile—with Russia and its historical predecessors. The Samis inhabited the Swedish mainland down to central Sweden, when the Scandinavians arrived. In Sweden, Sami has about 20,000 speakers, divided into several (five or more) languages. Some archaic Swedish varieties are occasionally held to be languages, e.g. Älvdalsmål in the province of Dalecarlia (Swe. *mål*, old form for *dialekt*). Capital Stockholm (metropolitan Stockholm area with 1.6 million inhabitants) is multilingual, with some suburbs almost exclusively inhabited by non-Swedish speakers.

The main migrant languages are Albanian, Arabic, Bosnian/Croatian/Serbian, English, Finnish/Meänkieli, Greek, Kurdish, Persian, Spanish, Turkish, and Polish. About 120 languages are taught in hour-based mother tongue instruction in the Swedish basic compulsory school. Since 2000, Sami, Finnish, Meänkieli, Romani Chib, and Yiddish are official minority languages.

Denmark (Danmark)

Denmark has a parliamentary and democratic monarchy, with a population of 5.3 million. Danish is the official language. It is the first language of about 4.9 million Danes. The capital Copenhagen area (København) has about 1.7 million inhabitants.

German is an official regional language in Southern Denmark (20,000 speakers in Northern Schleswig), and a notable population of speakers of Danish lives in Northern Germany/Schleswig, as a result of border treaties between Denmark and Germany. Denmark's long-term enemy was Sweden, with which Denmark fought about the dominance over the Baltic Sea area. Denmark lost three provinces on the Swedish mainland (Scania, Halland, and Blekinge) and Gotland to Sweden in the mid-seventeenth century. The Norwegian provinces of Härjedalen and Jämtland, then parts of Denmark, were lost to Sweden simultaneously. Danish is a former high variety and a secondary school language, in addition to Greenlandic in Greenland (an Eskimoic/Inuit language; *Kalaalit Nunat*) and to Faeroese in the Faeroes. The Faeroes received autonomy in 1948 and Greenland received autonomy in 1979. Danish was temporarily a high variety in Norway during the era of Danish dominance over Norway (1380–1814).

The main migrant languages are Turkish (50,000), German, English, the Scandinavian languages, Bosnian/Croatian/Serbian, Punjabi, Vietnamese, Persian, and Somali.

Norway (Norge, Noreg)

Norway has a parliamentary and democratic monarchy, autonomous since 1814 and independent since 1905, with a population of 4.6 million. The capital Oslo area has about 750,000 inhabitants. Two official varieties of Norwegian are spoken: Nynorsk and Bokmål (about 95% of all Norwegians speak one of them as their first language; Bokmål is chosen by 80-85%). Bokmål is mainly spoken in the urban and northern parts, and Nynorsk in the western/central rural regions. Of the 440 municipalities, 114 have declared themselves as officially Nynorsk-speaking, 175 have declared Bokmål as their official choice, and the rest, 151, are 'neutral'. The larger cities (Oslo, Bergen, Trondheim) have declared themselves 'neutral', but in practice there is a clear dominance for Bokmål.

The largest Nordic population of Sami speakers is found in Norway (30,000). Most of them belong to the North Sami language group. It is a regional official language in Finnmark and Tromsø provinces, Northern Norway, where some of the municipalities are bilingually administrated. Among the indigenous Kvens (who speak a Finnish-related language/variety), in and east of the town of Tromsø, about 3,000 know the language. Several main migrant groups are 'nonvisible minorities'. The main migrant languages are the Scandinavian languages (100,000), English (70,000), Punjabi (24,000), German, Bosnian/Croatian/Serbian, Vietnamese, Turkish, Persian, and Finnish.

Finland (Suomi)

Finland was a part of the Swedish kingdom for about 700 years, until 1809. It was the Autonomous Grand Duchy of the Russian Empire from 1809 to 1917, and became an independent republic in 1917. It has a population of 5.1 million, officially bilingual in Finnish and Swedish (nationally). Sami is an official regional language since 1991. Romani, Tatar, and sign language have been given minority status since 1994, and Russian is on the verge of becoming acknowledged

as one in 2001. In all, 93% speak Finnish as their first language. Finland Swedish (finlandssvenska) is spoken in the southern and southwestern coastal areas. The Finnish Åland Isles in the Baltic Sea have a specific legal and linguistic position; they are officially monolingual in Swedish (League of Nations and the UN, 1921). Finland Swedish, with 297,000 speakers (5.7%), is a fully protected official language (the Constitution and the Language Act, 1919, 1921 and 2004). Finland Swedish is referred to as the East Swedish dialect. It has an overt (Sweden Swedish) and a covert norm (Finland Swedish standard). It is also seen as an archaic Swedish variety, influenced by Finnish. Of the municipalities, 22 are officially Swedish-speaking (2001), 20 are bilingual in Finnish-Swedish, three are bilingual in Finnish-Sami, and 389 are monolingual in Finnish. The capital of Helsinki (Helsingfors; one million inhabitants in the metropolitan area) is bilingual. The vast majority of the Finland Swedish speakers are bilingual, who function as language and culture mediators from and to the other Scandinavian countries. Most Finns learn another national/'domestic' language in school, by force of the Educational Act. English competes with Swedish for Finnish-speaking pupils today. The main migrant languages are Russian (20,000 speakers), Estonian, Somali, Vietnamese, English, Kurdish, and Karelian.

Iceland (Island)

Iceland is a republic, which became independent in 1944, with a population of about 280,000. The capital area of Reykjavík has 150,000 inhabitants. Iceland was under Norwegian rule from 1380 to 1550, and under Danish rule from 1550 to 1848. Icelanders retained Icelandic as their *lingua sacra* during the Reformation. Icelandic is the official language. Iceland is one of the few practically monolingual nation-states of today. In school, Danish and English are secondary languages. An American airbase, Keflavik, is the only additional major language representation of Iceland, but it lives a separate life. Icelandic pupils learn an adapted variety of "Scandinavian" based on modifications of pronunciation and a common Scandinavian vocabulary, for their contacts with Scandinavians, since Icelandic differs considerably from the mainland Scandinavian languages.

Language and corpus planning play an active role in the everyday lives of Icelanders. Language contact effects are, however, inevitable, since quite a few Icelandic students temporarily study abroad (North America, Scandinavia). Today, modern media and communication technology also bring English into the everyday life of Icelanders.

Language Contact Types in Scandinavian/Nordic Past and Present

Several main types of language contacts have existed in Scandinavia. The oldest known one is that of the Sami and Scandinavian (in Sweden and Norway), and Sami and Finnish settlers (in Finland; 2500–200 BC). In older place names and special vocabulary, the Sami impact is still obvious. The impact of the majority languages is more apparent in Sami, however. High proportions of Sami speakers live in the three capitals: Stockholm, Oslo, and Helsinki.

The second type consists of Scandinavian expansionism, which includes the Viking era (800–1100). During this period, the Faeroes, Iceland, and Greenland received their Scandinavian settlers and later, dominance. The Scandinavian/Swedish settlement of southern and western Finland, a third type, is both older and simultaneous with this. These two types of migration formed the basis for Scandinavian "semicommunication" and later pan-Nordism (Swe. *nordism;* see below). Such receptive bilingual communication also developed due to the Danish dominance over Norway and Iceland. The changing borders between Sweden and Denmark–Norway reinforced this; when the former dialect areas of Denmark – Norway were Swedicized from the mid-1600s, new overlapping linguistic continua developed.

The fourth type is that of continuous migration and settlements of Finns in Sweden and Norway, from about the fifteenth century CE. Vast areas were settled and farmed by Finnish speakers. These so-called forest Finns in central Sweden and southeastern Norway have recently experienced language death. A separate migration was that of Finns to Stockholm, which has been continuous for about 800 years. Migration to Sweden was internal until 1809; later, it reemerged as emigration from Finland to Sweden. Fewer Finns have left for Norway and even less for Denmark. Migration in the other direction has been rare.

Fifth, Latin was the scholarly language for the mainland Scandinavian countries and Finland for several centuries, until the early nineteenth century. The former countries shifted to their national languages earlier than did Finland to Finnish, for cultural, public and educational purposes.

All of the countries faced extensive migration to North America, especially to the regions close to the Great Lakes, during the latter half of the nineteenth century. Scandinavians on an average moved earlier than Finns.

More recently, there have been sporadic migration waves to all of the Scandinavian countries, during and after World War II. Both labor migration and refugee waves have normally been punctual in kind, compared

to the continuous migration between the Nordic countries. Due to the booming post-World War II economy of the Nordic countries, first in Sweden and later in Norway, Denmark, and Finland, all countries have received notable numbers of migrants from first, the neighboring countries, then from Southern Europe, and more recently globally. The most significant of these migration waves was that of Finns to Sweden from the late 1950s to the late 1980s. The free Nordic labor market was one important contributing factor. Today, there are groups of equal size of speakers of Swedish in Finland, as there are speakers of Finnish in Sweden.

A last type of language contact is a matter of cultural and linguistic transfer, rather than the transfer of people. Earlier, Latin, German, and French played such roles. The increasing presence of English in Scandinavia has been seen as an asset since the 1950s, especially in Sweden, but has recently become a matter of concern. English is diminishing the language use domains of the national languages. English is by far the most popular foreign language in school, and it is compulsory. Its impact differs from earlier foreign languages, which were transplanted for specific purposes or by socially defined groups. English is now used as a language of wider communication, and English expressions occur in all types of domains and in all social groups.

The common linguistic and cultural roots, in combination with the later separate historical and political developments in the three countries, make the problems of the concept of language evident. If languages are defined as language forms characterized by mutual intelligibility, Swedish and Norwegian or Norwegian and Danish could be defined as dialects rather than separate languages. On the other hand, speakers of some of the Swedish dialects, such as Scanian (skånska) in the south of Sweden, which until 1645 belonged to Denmark, the northernmost Swedish dialects, which bear evidence of language contacts with Sami and Finnish, or the archaic Gutnish dialects of Gotland, do not share full intelligibility with all other dialects.

Scandinavian Expansionism and Retreat

A division into West vs. East Scandinavian started in Old Scandinavian in the sixth or seventh century, when Norse farmers started sailing westward and inhabited the Faeroes and Iceland. The process was completed in the ninth and tenth centuries CE. This contiguous migration also reached Greenland. From Greenland, Leif Eriksson sailed westward and "discovered" North America in the eleventh century CE. New varieties of Old Norse developed. The daughter languages Icelandic and Faeroese developed, in addition to two extinct languages, Orkney and Shetland

Norn. Initially, speakers of Insular Nordic languages presumably had contacts with Celtic speakers, who had inhabited the same areas temporarily.

Scandinavian raids and trade created external language islands in vast areas of Northern and Western Europe. The Viking age (c. 800–1100 CE) started with tribal chieftains' small-scale raiding. Later, this turned into a large-scale warfare led by powerful Viking dynasties. The improved shipbuilding techniques contributed to this. The Viking period thus changed the language history of e.g. the British Isles. A landmark was the Danish conquest of Britain in the eleventh century CE. The Danes later continued their expansion in more peaceful ways. The Norwegians concentrated their raids to the islands north and west of Scotland and England, including Ireland, Isle of Man, and Scotland itself. Dublin was founded by a Norwegian chieftain in 841 CE. In Normandy, a Danish Duchy was founded in the tenth century by King Rollon. The Swedes established their rule in Kiev, Ukraine, before the Viking age and later in the Baltic countries. Russia was a target for early Swedish Viking trade and raids. Later, it became the target of the unified Swedish kingdom's attempts both to expand and protect itself. Finland was Christianized by the Swedes and thus came under its religious and cultural impact from the twelfth century, even if Scandinavians had lived in the coastal areas for centuries by then.

When the dominance of Denmark weakened, the Viking rule in the Baltic Sea was soon replaced by that of others. The commercial union of the Hanseatic League, especially under the leadership of Lübeck in the fourteenth and fifteenth centuries, was particularly successful. For some time, the League ruled the whole Baltic Sea and the North Sea. The establishment of Low German merchant layers in the main cities, Stockholm, Kalmar and Visby in Sweden, Copenhagen in Denmark, and Bergen in Norway, paved the way for an extensive linguistic influence on the local languages. In Sweden, this created a new prestige language; Low German became the language of the merchants, and the townspeople became highly influenced by its cultural impact. About one-third of modern Swedish consists of Low German vocabulary, for example, *släkt* 'family, relatives', *handel* 'trade', *krig* 'war', *möjlig* 'possible', *bliva* 'be, become', *smaka* 'taste', *skriva* 'write', and *men* 'but'. This indicates a high degree of bilingualism among important sections of Swedish speakers. Another central contribution from Low German was the borrowing of derivational prefixes, like *an-, be-, för-: använda* 'use', *behöva* 'need', *förstå* 'understand', and suffices like *-era, -het, -inna: studera* 'study', *svaghet* 'weakness', *lärarinna* 'female teacher'.

Scandinavian 'semicommunication' refers to the possibility to use one's own language, and to understand

and be understood by speakers of other Scandinavian languages. It functions thanks to the common background of the Scandinavian languages. Finns have in addition learned some Swedish in school or needed it for their civil careers, for several centuries. Icelanders have learned both Danish and, more recently, adapted 'Scandinavian'. With some minor phonological modifications, the use of common vocabulary, and avoidance of 'difficult' language-specific features, most Nordic citizens knowing a Scandinavian language may communicate among themselves. The demand to modify speech is the highest for Danes and Icelanders, who are the furthest away from modern, common Scandinavian. The common historical basis of Danish, Norwegian, Icelandic, and Swedish is one prerequisite of this type of communication. Another is the idea and philosophy of a shared political agenda, which to a high degree was created in the latter half of the 1800s. This pan-Nordic idea means that the countries do not only share historical and cultural features, but they have also attempted to teach each other's languages and about each other's countries in school. Even today, Nordic people may present a common face to other countries and in international cooperation, based on shared values. During the last decades, this linguistic understanding and communicative potential have weakened, and so has the striving for a common political agenda. Denmark, Finland, and Sweden are members of the European Union (EU); Norway and Iceland are not. Nevertheless, in practical administrative, political life and in the attitudes of people from the Nordic countries, there is a sense of togetherness and regional identity, which goes beyond ordinary neighborhood feelings. The Samis were more or less excluded from such cooperation earlier, but they have now created their own Nordic networks, both between Samis in different countries, and between different Sami groups in the same country.

The original Nordism of the 1860s has later been extended in many formalized ways. The exchange of ideas, culture, and ordinary people between the countries has continued and enabled the historically based Scandinavian communication to be prolonged in the form of mutual understanding of each other's languages. Today, Nordism is getting competition from EU- level cooperation. To some extent, it is also challenged by the attempts to integrate the three Baltic countries, Estonia, Latvia, and Lithuania, into the Nordic sphere, in which cooperation in English is preferred.

References

Aikio, Samuli, et al. 1994. The Sami culture in Finland. Helsinki: Lapin Sivistysseura.

Att förstå varandra i Norden, (To understand each other in the Nordic countries). Stockholm: Nordiska rådet/Nordiska ministerrådet/Nordiska Språkrådet, 1997.

Blom, Gunilla, et al. (ed.) 1992. Minority languages: the Scandinavian experience. Oslo: Nordic Language Secretariat.

Börestam Uhlmann, Ulla. 1994. Skandinaver samtalar, (Scandinavians in conversation). Skrifter från institutionen för nordiska språk, 38. Uppsala: Institutionen för nordiska språk vid Uppsala universitet.

Hasselmo, Nils. 1974. Amerikasvenska (American Swedish). Skrifter utgivna av Svenska språknämnden, 51. Lund. Stockholm: Svenska språknämnden.

Haugen, Einar. 1969. The Norwegian language in America: a study in bilingual behavior, Vols. I–II, 2nd edition. Bloomington: Indiana University Press.

Melander, Björn (ed.) 2000. Svenskan som EU-språk (Swedish as a EU Language). Uppsala: Hallgren & Fallgren.

Simonsen, Dag F. 1996. Nordens språk i EU:s Europa (The languages of the Nordic countries in the Europe of the European Union), Rapport 22. Oslo: Nordisk Språksekretariat.

Vikør, Lars. 1993. The Nordic languages: their status and interrelations. Oslo: Novus Press.

Virtaranta, Pertti, et al. 1993. Amerikansuomi. (American Finnish.) Tietolipas 125. Helsinki: Suomalaisen Kirjallisuuden Seura.

Web-sites:

www.dst.dk (Statistics Denmark)
www.ssb.no (Statistics Norway)
www.scb.se (Statistics Sweden)

JARMO LAINIO

See also **Swedish and Scandinavian languages; Finnish and Finnic Languages; Language: Contact—Overview; Bilingualism; Language Planning**

Searle, John R.

John Searle, the contemporary analytic philosopher, rose to fame with the publication in 1969 of *Speech acts: An essay in the philosophy of language* (SA). In this book, Searle systematized a number of ideas that were originally developed by J.L. Austin whom he knew as a graduate student at Oxford in the 1950s.

Searle's reputation as Austin's main intellectual heir, while no doubt right as far as it goes, must nonetheless be properly qualified. On the one hand, it should not be allowed to eclipse his own important contribution to not only the theory of speech acts in particular and philosophy of language in general but also to the philosophy of mind, a branch of philosophy in which he has established himself as a key, although contentious, figure. On the other hand, it is not entirely true either that his chief merit consisted in simply systematizing and bringing to fruition what his teacher Austin would have done all by himself, had an untimely death not prevented him from doing it. Unfortunately, many writers have helped create the widespread misperception that Searle simply took over from where Austin left off or, equivalently, that, except for some minor differences here and there, the work of the two philosophers constitutes a smooth continuum.

Searle's originality as a philosopher was announced on the very first pages of SA, where he insisted that his work was, as the subtitle proudly proclaimed, an exercise in the philosophy of language rather than Austin's own narrower 'linguistic philosophy'. In fact, Searle has ever since shown little interest in some of the major tenets of linguistic philosophy, notably the idea that a careful analysis of the ordinary, everyday use of language can help solve a great number of philosophical puzzles. Searle's central concern was announced in SA in the form of two aphorisms: (a) Speaking a language is engaging in a form of rule-governed behavior.(b) The basic unit of linguistic communication is the speech act. Searle also made his differences with Austin very clear in his early essay 'Locutionary and illocutionary acts', wherein he argued that Austin's concept of a 'locutinary act' should be replaced by his own notion of a 'propositional act'. For Austin, locutionary acts were the acts of producing utterances with a certain sense and meaning and these were opposed to illocutionary acts (performing certain acts such as promising *in* saying certain words) and perlocutionary acts (doing certain things such as threatening *by* saying certain words). By introducing the concept of proposition into speech act theory, Searle put Austin's philosophical reflections back on the beaten track of analytic philosophy. He also prepared the grounds for locating certain universals in speech acts across the multitude of world's languages, a quest that Austin would in all likelihood have dismissed as premature at best and a wild goose chase at worst. For Searle, illocutionary acts had the form 'F (p)' as their canonical structure—where 'F' denoted an Illocutionary Force Indicating Device (IFID) and the 'p' represented the 'propositional content' of the act in question. As many scholars have pointed out, Searle gave a distinctive Platonic twist to ideas he had origi-

nally picked up from his Oxford mentor who was self-declaredly an Aristotelian.

There can be little doubt that the theory of speech acts, in the form in which it is known to scholars outside the disciplinary bounds of analytic philosophy, is largely due to Searle's decisive intervention. This is especially the case in areas such as linguistics, psychology, sociology, and anthropology where speech act theory has had a tremendous impact. In the early 1970s, for instance, there was a very strong dissident movement called Generative Semantics within the Generative paradigm in linguistics that, spurred on by Searle's innovative idea of prepositional acts, sought to incorporate the concept of speech acts into the semantic component of grammar by stipulating that every sentence had an underlying form with a highest clause that specified its speech act status.

With the publication of *Expression and meaning* (EM) in 1979, Searle marked the consolidation of a theory that, after two decades of rapid expansion and intensive research, had important things to say about a range of interesting phenomena that included metaphor, fiction, and the so-called indirect speech acts (such as, say, the performance of an act of request by asking one's interlocutor a question—as in saying 'Can you please pass the salt' to someone sitting next to you at a dinner table). Searle also indicated that his own future work in the philosophy of language was going to be strongly influenced by the concept of intentionality. In fact, along with his interest in bringing the speech act theory in line with the thesis of intentionality, Searle also announced (in the preface to EM) yet another task for the years ahead: that of providing an adequate formalization of the theory using the resources of modern logic.

In 1983, Searle published *Intentionality* and two years later *Foundations of illocutionary logic* (FIL), in collaboration with the Canadian logician Daniel Vanderveken. Subtitled *An essay in the philosophy of mind*, *Intentionality* marked a decisive turn in Searle's philosophical enterprise and prepared the ground for his later distinctive and often polemical interventions in the then burgeoning new discipline called 'cognitive science'. As regards the second research direction announced in EM, FIL did fulfill the promise. But the scholarly community has been rather critical of the line of inquiry announced in FIL and, to judge from Searle's own lack of enthusiasm in pursuing it, the program no longer seems to enjoy the support of the Berkeley philosopher (although his erstwhile collaborator Vanderveken has single-handedly carried the torch since then).

Scholars are also divided in their full support for the turn marked by *Intentionality*, especially insofar as it is meant to be a complement to his earlier work on speech

acts. Many resent what they see as an exaggerated mentalist swing in conceptualizing speech acts, accompanied by Searle's interest in looking for a universal base to what many see as acts tethered to specific cultures. Some, notably certain scholars with an ethnomethodological orientation, regard Searle's mentalism as a step in the wrong direction away from Austin's more culturally sensitive approach to language.

From the early 1980s onward, Searle has been very active in the philosophy of mind. He has been a tireless critic of those cognitive scientists who have equated the working of a human brain with that of a digital computer. In the philosophy of mind, Searle is also known for his work on the problem of consciousness. His famous Chinese Room Argument, put forward in 1980, according to which a mere computer model of associating symbols is no guarantee of true knowledge of the language in question (say, Chinese), has been discussed over and over again in the literature and has influenced discussion in the area ever since.

Besides his intensive work in academic philosophy, Searle has been an active contributor to periodicals such as the *New York Review of Books,* where he has published scores of reviews and write-ups sharply critical of such contemporary trends as poststructuralism, pragmatism, and deconstruction. He has also taken highly polemical positions on student politics, and politicization of life on campuses.

Biography

John R. Searle has been Mills Professor of the Philosophy of Mind and Language at the University of California at Berkeley since 1959. He was born in Denver, Colorado (USA) in 1932. His father was an AT&T executive and his mother a medical doctor. As a child, he moved to New York, then to New Jersey, and finally to Wisconsin, where he graduated from high school. He began his university work at the University of Wisconsin in 1947, where he stayed till 1952. At 19, as a Rhodes Scholar, he went to England and studied under P.F. Strawson and J.L. Austin at Oxford University (1952–1959), where he took his B.A. First Class Honours (1955), M.A. (1959) and D.Phil. (1959). He was a Lecturer in Philosophy at Christ Church (Oxford) (1957–1959). In 1969, he published *Speech acts: An essay in the philosophy of language*. The same year, he took up a teaching position at UC Berkeley. Searle was elected for Distinguished Teaching Award by UC Berkeley (1999) and was awarded Honorary Degrees by the Universities of Wisconsin (1994) and Adelphi (1993), among others. He has also been a visiting professor at many universities across the five continents.

References

Burkhardt, A. (ed.) 1990. Speech acts, meaning and intentions. Berlin: Walter de Gruyter.

Lepore, E., and R. Van Gulik (eds.) 1991. John Searle and his critics. Oxford. Blackwell.

Love, N. 2000. On Searle on language. Language & Communication 19.

Parret, H., and J. Vershueren (eds.) 1992. (On) Searle on conversation. Amsterdam: John Benjamins Co.

Rajagopalan, K. 2000. On Searle [on Austin] on meaning. Language & Communication 20.

Searle, John R. Speech acts: an essay in the philosophy of language. New York: Cambridge University Press. 1969.

———. 1971. The campus war, New York: The World Publishing Co.

———. 1979. Expression and meaning: studies in the theory of speech acts. New York: Cambridge University Press.

———. 1984. Minds, brains and science. London: Penguin Books.

———. 1983. Intentionality: an essay in the philosophy of mind. New York: Cambridge University Press.

———. 1992. The rediscovery of the mind. Cambridge, MA: MIT Press.

———. 1995. The construction of social reality. London: Penguin.

———. 1997. The mystery of consciousness. New York: New York Review Press.

———. 1998. Mind, language and society, philosophy in the real world.

Tsohatzidis, S.L. (ed.) 1994. Foundations of speech act theory. London: Routledge.

KANAVILLIL RAJAGOPALAN

See also **Austin, John Langshaw; Speech Acts**

Second Language: Acquisition

It has been said that multilingual human brains are more common than monolingual ones; most of the people on this planet know more than one language. A major insight from the field of second-language acquisition (SLA) has been that the grammars of second-language learners are systematic and rule-governed.

These so-called 'interlanguage' grammars contain elements from the first language (L1)—e.g. German final devoicing or French verb raising—and elements from the second language (L2)—e.g. Canadian vowel raising or Spanish null subjects. A learner's production (even if it includes errors) is the product of the mental representation of the grammar. There has been debate as to whether interlanguage grammars are subject to the same constraints as the grammars of primary languages—the 'Access to UG' debate. Related to this is the question of whether L2 learners can acquire a linguistic structure (e.g. gender, tone, [strident]) if that structure is completely absent from the L1.

A basic insight from cognitive theory that helps us to understand SLA involves knowledge vs. skill. The knowledge that we store in our heads is a relatively stable trait. You either know the word 'cat' or you do not. You either know that the sentence, 'They is inable to speaked French' is ungrammatical or you do not. Clearly, you have to acquire knowledge of your L2, but you also have to acquire skills. You have to be able to comprehend fast speech, or carry on a conversation. Proficiency in a second language is a complex construct that includes a range of knowledge (from grapheme to phoneme to sentence to text) and a range of abilities (from politeness routines to appropriate register). Can adults attain nativelike proficiency in a second language? Often, this question is investigated under the rubric of the critical period hypothesis. There are undeniably age-related effects in L2 learning; people who start acquiring their L2 early are less likely to have an accent. However, as we have seen, accent is a small part of L2 proficiency. Is there some point after which nativelike proficiency is impossible? The answer to this question depends on which elements of proficiency one is looking at. White and Genesee (1996) have argued that nonnative speakers who started their L2 learning later in their lives can evidence grammatical knowledge and performance that is statistically indistinguishable from native speakers. Bongaerts (1999) has shown this for nonnative accents as well.

One of the most recognizable characteristics of adult SLA is that a great many of the people who attempt it do not sound like native speakers of the language; they may have an accent. This has led some to argue that SLA is fundamentally different from first-language acquisition in that the end point (or ultimate attainment) is so divergent. At the sound level, L2 learners can sound nonnativelike for two reasons. Let us take an example from stress. Many languages mark certain syllables as prominent by stressing them. The second syllable in 'banána' is more prominent than the first or third syllables. Languages vary in where they put the stress, so it is something that has to be learned.

Therefore, this is one possible source of error, and hence one way in which people can sound nonnative-like; they can get the stress placement wrong. This would be a phonological problem. They could, however, put the stress on the right syllable but still not sound like a native speaker, Maybe in their L1, they indicate stress by loudness or vowel length, whereas in English the main means of indicating stress is by a pitch change. This would be an example of a phonetic marker of their accent. The question of why some sounds are harder to learn than others has been addressed from typological (Eckman), phonological (Brown), and phonetic perspectives (Flege).

Languages also vary in terms of their syntactic structures. In English, you need to have a subject for your sentences. You can say 'She speaks French', but not '*Speaks French'. In languages like Spanish, though, you can drop your pronominal subject and say either 'Yo hablo español' or 'Hablo español'. Hence, Spanish speakers of English are going to have to learn that they cannot drop their subjects in English. L2 learners will also have to acquire the grammatical structures necessary for making questions, forming passives, or embedding clauses (just to name a few). They will have to learn to interpret quantifier scope and other subtle semantic properties. These particular structures can be used to identify two main approaches to explanation in SLA.

There are some who argue that structures that are common (or natural) in the world's languages will be easy to learn (regardless of your L1). For example, every language in the world that has null subjects also allows sentences with pronominal subjects. The presence of null subjects implies the presence of subjects. There are no languages that allow *only* null subjects. In this framework, then, it is assumed that null subjects are more marked than pronominal subjects. Furthermore, it is assumed that structures that are more marked will be more difficult to acquire (Eckman 1991).

Generative linguists have also been interested in explaining L2 acquisition. Consider the above examples. Linguists might invoke learnability issues to explain these facts. An English speaker trying to learn Spanish will hear Spanish sentences that have null subjects; there will be positive evidence in the linguistic environment that the learner's current grammar (or hypothesis) is wrong, and they must change it. The Spanish speaker trying to learn English, on the other hand, will begin the process with a grammar that allows both null and overt subjects. When listening to the English sentences of the linguistic environment, they will never hear any sentences that are inconsistent with their grammar. There is no reason for them to change. It has been argued that this may explain some Directionality of Difficulty effects observed in SLA.

939

English speakers have little difficulty in learning to drop their pronouns in Spanish, while Spanish speakers can have trouble learning to always provide a pronominal subject in English.

References

Archibald, J. (ed.) 2000. Second language acquisition and linguistic theory. Oxford: Blackwell.

Bley-Vroman, R. 1989. What is the logical problem of foreign language learning? Linguistic perspectives on second language acquisition, ed. by S. Gass and J. Schachter, 41–68. Cambrdige: Cambridge University Press.

Bongaerts, T. 1999. Ultimate attainment in L2 pronunciation: the case of very advanced late L2 learners. Second language acquisition and the critical period hypothesis, ed. by D. Birdsond, 133–159. Mahwah, NJ: Lawrence Erlbaum and Associates.

Brown, C. 2000. The interrelationship between speech perception and phonological acquisition from infant to adult. In Archibald, pp. 4–63.

Eckman, F. 1991. The structural conformity hypothesis and the acquisition of consonant clusters in the interlangauge of ESL learners. Studies in Second Language Acquisition 13(1). 23–42.

Flege, J. E. 1992. Speech learning in a second language. Phonological development, ed. by C. Ferguson, L. Menn, and C. Stoel-Gammon, 565–604. Timonium: York Press.

Flynn, S., G. Marthohardjono, and W. O'Neil. 1998. The generative study of second language acquisition. Mahwah, NJ: Lawrence Erbaum and Associates.

White, L. 1989. Universal grammar and second language acquisition. Amsterdam: John Benjamins.

White, L., and F. Genesee. 1996. How native is near-native? The issue of ultimate attainment in adult second language acquisition. Second Language Research 12(3). 233–265.

JOHN ARCHIBALD

Second Language: Learning

The ability to learn another communication system, subsequent to a first, acquired during infancy and early childhood, is unique to the human species. In their use of second languages, humans easily outperform even those species endowed with highly elaborate communication systems, among them apes trained to mimic human language. Yet, the question of whether it is possible to learn a second language just like the first, i.e. to proceed through similar developmental phases and attain the same type of knowledge and skills, is controversial. Answers to this and related questions promise to shed light on the nature of the human language capacity, which may explain why much of the research on second-language learning addresses problems that involve comparisons between first- and second-language acquisition.

First languages (L1) are generally understood to be the end products of developments guided by the human language faculty. The core properties of its grammatical components are captured in terms of principles of Universal Grammar (UG). Some of these principles apply invariantly to all languages, whereas others are 'parameterized', i.e. they offer two (or more) options ('parameter values') in which languages may differ. An invariant principle specifies, for example, that sentences need to contain a subject position; a parameterized option allows for this position to be lexically empty (e.g. Spanish—*trabajo en casa*) or it needs to be lexically filled (e.g. its English equivalent *I work at home*). According to this view, UG is a theory about the initial state of language development in that it specifies the kind of implicit, innate knowledge the child brings to the task of acquiring a first language, preceding experience. Moreover, UG constrains the developing grammatical systems so that at every point of development, children's immature grammars UG-conform and are thus 'possible human grammars'. Consequently, L1 acquisition consists of both learning from experience (e.g. lexical learning, learning of language-specific properties) and the activation ('triggering') of principles of UG. When languages vary, children are said to set parameters on one of the given options ('values'), based on their experience with target input. Children growing up with two or more languages can set a parameter on different options, if required by the target systems. Bilingual acquisition during childhood is thus an instance of simultaneous acquisition of two 'first' languages.

Second languages (L2), on the other hand, result from learning processes initiated once a first language has already been acquired, either fully, in the case of adult L2, or partially, in the case of child L2 acquisition. Distinguishing between various acquisitional types implies that substantive differences exist between L1 and L2, and less important ones between child and adult L2. Although much research supports this view, the issue is far from settled. Particularly, the critical age range separating L1 from L2 is a matter of much

controversy; it is clearly not the same for phonological, syntactic, and morphological properties of the target language, with the ability to acquire a native phonology fading out first, possibly at around age 6. A conservative hypothesis consists in classifying as instances of 'L1' those cases in which the child has been exposed to one or more languages before age 3, as 'child L2' cases in which the onset of learning of one or more additional languages lies between the age range 5 and 8 years of age, and as 'adult L2' those cases in which the onset of learning lies after 10. Further research is needed in order to fill the gaps in the given age ranges and to decide whether the critical distinction is indeed the one between L1 and child L2 acquisition.

Descriptively, L1 and L2 acquisition differ in important aspects, namely in the course of acquisition as well as in the final state attained by learners. As for explanation, the question is whether L2 acquisition is guided by the human language faculty and, consequently, whether it eventually leads to a kind of knowledge of the target language that is identical to or at least resembles that of native speakers of the language. This amounts to asking to what extent L2 acquisition is determined by principles of UG. If the approximative systems underlying L2 learners' language use should turn out not to be constrained by UG principles, it would be questionable to refer to them as to 'grammars', in the sense defined for L1. Second languages might then possibly be defined not as 'natural languages', contrary to our initial assumption.

In at least one regard, L2 acquisition does indeed resemble L1 development: one finds strictly ordered developmental sequences. Many, although not all, grammatical characteristics of a language emerge in a fixed order across learners, independently of varying learning contexts, individual properties of the learners, or structural features of their first languages. In the 1970s, the 'morpheme order' studies showed that certain grammatical elements like verb endings (-s, -ed, -ing, etc) tend to appear in the same order in the speech of different child and adult L2 learners of English. Subsequent research analyzing a larger variety of grammatical phenomena (particularly word order) in a number of languages revealed that structural properties of the target languages emerge crossindividually in an invariant order, although the later ones do not depend, in principle, on the earlier ones. Moreover, the order of acquisition and use found in naturalistic L2 acquisition proved to be difficult to manipulate by means of foreign language instruction. It seems, thus, that there exists a hidden logic by which learners of a given target language are guided. A major goal of L2 research is, therefore, to uncover the underlying logic, for this should be crucial in deciding if second languages are indeed natural languages. In view of the fact that L2

acquisition proceeds through an ordered sequence of stages, it is tempting to conclude that both L1 and L2 are guided by the same underlying principles. However, differences also manifest themselves with respect to the developmental sequences. The morpheme order studies showed that the sequences are not exactly the same in L1 and in L2 acquisition, a finding confirmed by subsequent research for a broad range of grammatical phenomena. In sum, although both types of acquisition exhibit a course of development that is uniform across individuals, these invariant aspects are not the same in L1 and L2 acquisition, thus raising the question of whether the underlying principles and mechanisms are identical.

In L1 acquisition, developmental sequences are commonly explained through UG, i.e. the succession of L1 developmental phases can be understood as a reflection of the fact that children increasingly gain more complete access to, or become able to use, the principles of UG and set parameters for the values required by the target grammar. Prerequisites for this development are, most importantly, a growing lexicon and the ability to deal with longer and more complex constructions, as a result of a more powerful working memory and increasing processing capacities.

This kind of explanation cannot simply be carried over to L2 acquisition; older learners can rely on fully developed working memories and on mature language processing systems. While explaining L1–L2 differences, it also suggests that adult L2 learners should outperform child L2 and L1 learners—hardly a satisfying result in view of the fact that L1 development in nonpathological cases leads to remarkably uniform results (despite considerable differences in learning environments), whereas this is only true, at best, for a small group of L2 learners. A possible solution for this problem is that mechanisms of language use play different roles in L1 and L2 acquisition. Whereas in all types of acquisition, they influence the choice of constructions, resulting in preferences of use for particular forms, in L2, they might determine the internal order of acquisitional sequences. This proposal is based on the insight that processing of linguistic forms and constructions requires a certain cognitive capacity. Given that the total amount is limited, learners need to keep low the cognitive costs required by formal aspects of language in order to deal with problems of content and with communicative tasks. Consequently, they first acquire those constructions requiring the least amount of processing capacity. If correct, processability helps to determine the underlying logic of L2 acquisitional sequences. In L1, they follow essentially a grammatical agenda.

Assuming that the explanation of L2 sequences in terms of processability of linguistic forms is correct

does not entail that the human language faculty has no role to play in L2 acquisition. A plausible scenario predicts that the various cognitive resources compete, and that it is due to the fact that some of them, e.g. memory and processing, are more developed that they interfere with the grammatical module in a way that is not possible during early childhood in L1 development. The well-foundedness of this competition hypothesis is difficult to assess, however, since not much empirical work has been dedicated to its scrutiny. Instead, much research since the mid-1980s has tried to find empirical evidence to show that UG is still accessible in L2 acquisition. Further possible sources of knowledge for L2 learners are their respective first and other previously learned languages. To the extent that this type of knowledge interferes with what is offered by UG, this might contribute to an explanation of L1–L2 differences.

The contribution of principles of UG, on the one hand, and transfer of previously acquired knowledge, on the other, can be illustrated by considering the initial state of the learning process. In L1, the initial state is determined by the human language faculty and thus by UG. As for L2, there is broad consensus that learners, at the onset of acquisition, make use of knowledge sources not available to the L1 child, for example, grammatical knowledge or mechanisms of language use previously acquired. The exact nature and amount of transfer, however, is a particularly controversial issue. In the vast amount of publications dedicated to this topic over the last 50 years of research, radically different hypotheses have been defended, from 'full transfer' to attributing a marginal role to syntactic and morphological transfer. It is accepted that previous linguistic experience does influence L2 acquisition, resulting in a different approach to the learning task by L2 as compared to L1 learners. A frequently postulated, although contested, hypothesis claims that parameter settings are transferred from the L1 grammar to the L2 competence. Consequently, learners need to make appropriate changes in order to attain target-conforming knowledge of their L2 system, if source and target grammars differ with respect to particular parameter settings. In principle, this can be achieved either by inductive learning or by 'resetting' the parameters in question. Parameter 'resetting', however, presupposes that learners continue to have access to the various parameter values provided by UG, and this issue therefore represents the core problem in the debate about continuous access to UG. Whereas some researchers claim that L2 learners have 'full access' to UG principles and parameters, others argue that direct access is ruled out for principled reasons and that UG shapes the L2 system merely via the L1 grammar. An intermediate position is defended by those who favor the idea of partial access to UG, meaning that only nonparameterized

principles of UG are operative in L2 acquisition, whereas parameter values not instantiated in the L1 grammar are not available any more, and the respective properties of the target grammars need to be learned empirically. Although some evidence supports the latter hypothesis, the controversy is far from settled and this issue is likely to occupy L2 research in the future.

A further important source of knowledge, in addition to UG and transfer, is the learners' linguistic environment, i.e. input and feedback offered by their interlocutors. Their fully developed cognitive capacities and communicative skills may be expected to enable them to make better use of this source than L1 children. Numerous studies indeed confirm that native interlocutors tend to modify their language use when addressing L2 learners who, themselves, can actively engage in negotiations to provide them with a possibly more comprehensible input, corrections, etc. All this happens in naturalistic acquisition, which has been the object of the preceding remarks since there is no evidence that acquisitional processes differ significantly in tutored learning. But addressee-adapted speech, especially rejection and correction (negative feedback) of target-deviant learner utterances, is of special importance in classroom learning. The question, however, of whether these and other types of input modifications have substantial and long-term effects for the course and the result of learning processes is difficult to answer. From all we know, developmental sequences are not altered, but the rate of acquisition can be speeded up; moreover, certain acquisitional tasks (phonological and lexical learning) seem to be easier to influence in this way than others (grammatical morphology and syntax).

A characteristic feature of L2 learning is its interindividual variability. In spite of the fact that one finds invariable developmental sequences, individuals and groups of learners differ considerably. This refers not only to acquisition rates, use of transfer, etc. but, most significantly, also to ultimate success. Undoubtedly, some learners acquire the knowledge and skills enabling them to perform in a manner that makes their language production indistinguishable from that of native speakers. But we have no satisfying answer, as yet, to the question of why only a few learners achieve nativelike grammatical knowledge. A possible solution is that differences across learners are due to sociopsychological factors, most notably their attitudes and motivation toward the learning task, the target language, etc. But although this is certainly a plausible hypothesis, evidence demonstrating a cause-and-effect relationship between such factors and particular types of learning is scarce.

In conclusion, second-language learning, whether in a naturalistic or in a classroom setting, proceeds

through acquisitional sequences, largely independent of individual factors or contextual influences. These sequences are not identical, however, to those found in L1 development. Moreover, L2 learners vary considerably in the kind of use they make of their linguistic knowledge and in their ultimate attainment. For the time being, only partial and temporary answers are possible to the questions of whether L2 acquisition is guided by UG principles and whether second languages are 'natural languages'. It is not even clear whether 'perfect learners' have acquired a competence of the same nature as the grammatical knowledge of native speakers. Overall, the observable L1–L2 differences suggest that the underlying systems exhibit qualitative differences. This is supported by neurolinguistic evidence, which indicates that grammatical information from second languages is, at least partially, processed differently and in different areas of the brain, when compared to first languages. One may thus speculate that L2 knowledge should be characterized as a hybrid system, combining grammatical principles and mechanisms constrained by the human language faculty with others learned by means of general problem-solving capacities of the cognitive system.

References

Bley-Vroman, Robert. 1989. What is the logical problem of foreign language Learning? Linguistic perspectives on second language acquisition, ed. by Susan M. Gass and Jacquelyn Schachter, 41–68. Cambridge: Cambridge University Press.

Carroll, Susanne E. 2000. Input and evidence: the raw material of second language acquisition. Amsterdam: John Benjamins.

Clahsen, Harald, and Pieter Muysken. 1986. The availability of universal grammar to adult and child learners: a study of the acquisition of German word order. Second Language Research 2(2). 93–119.

Dulay, Heidi C., Marina K. Burt, and Stephen Krashen. 1982. Language two. Oxford: Oxford University Press.

Felix, Sascha W. 1987. Cognition and language growth. Dordrecht: Foris.

Long, Michael H. 1990. Maturational constraints on language development. Studies in Second Language Acquisition 12(3). 251–85.

Meisel, Jürgen M., Harald Clahsen, and Manfred Pienemann. 1981. On determining developmental stages in natural second language acquisition. Studies in Second Language Acquisition 3(2). 109–35.

Pienemann, Manfred. 1998. Language processing and second language development: processability theory. Amsterdam: John Benjamins.

Ritchie, William C., and Tej K. Bhatia (eds.) 1996. Handbook of second language acquisition, San Diego, CA: Academic Press.

Schwartz, Bonnie D., and Rex A. Sprouse. 1996. L2 cognitive states and the full transfer/full access model. Second Language Research 12(1). 40–72.

White, Lydia. 1989. Universal grammar and second language acquisition. Amsterdam: John Benjamins.

Wode, Henning. 1981. Learning a Second Language: An Integrated View of Language Acquisition. Tübingen: Narr.

JÜRGEN M. MEISEL

Second Language: Teaching

Second-language teaching (SLT) involves teaching one language to speakers of another language. A distinction is often made between SLT and foreign-language teaching (FLT). The latter refers to teaching a language not widely used in the community in which the language is being learned. For example, teaching English to students in Italy or Japan is often termed FLT, whereas teaching English to learners in the United States and in countries such as Nigeria or India, where English is an official language, is usually referred to as SLT.

For both SLT and FLT, several issues have to be addressed. These include the purpose for which the target language is taught (aim), what is taught (syllabus), how the content is taught (methodology), and how learning is assessed (evaluation).

Consideration of the learners' background is very important when considering the aim of the program. Key factors include the learners' age and proficiency in the target language; their linguistic, cultural, and educational experience; the time and money available for instruction and self-study and associated learning resources; and their reasons for learning the language. Some of these purposes may be quite general. Others may be very specific: for example, airline pilots need English for occupational purposes, British students wishing to attend university in Paris may study French for academic purposes, or Japanese engineers may need to read scientific and technological reports written in English. The analysis of backgrounds, needs, and resources facilitates decisions about the composition of classes, the length and frequency of instruction, the type of teacher deemed most suitable, and the syllabus.

Most SLT syllabuses in Britain and the United States 40 years ago were structural in orientation—that

is, they were based on the formal elements of phonology, syntax, and morphology in the language. Learners were guided to master the formal content of the syllabus at one level of difficulty before moving on to the next. Many SLT programs are still designed in this way. Since the 1970s, however, a wider view of the authentic use of language and increasing attention to the needs of learners have led to the adoption of alternative approaches to syllabus design—especially regarding the teaching of English. Predominant among these is communicative language teaching (CLT), which sees the communication of information and ideas as the primary goal of second-language learning. The emphasis is less on what learners know about the target language at the sentence level (linguistic competence) and more on what they can do with the language in longer stretches of spoken or written discourse (communicative competence). Instead of a linear progression of items presented in order of structural difficulty, language points are sequenced in terms of their perceived usefulness, and they are integrated with other elements and constantly revisited. In the 1980s, there were 'strong' and 'weak' CLT syllabuses; proponents of the strongest versions argued that the explicit teaching of grammatical structures hinders the acquisition and communication of meaning. Weaker versions of CLT incorporated structural elements within a broader framework of communicative functions, language skill development, and interactional learning strategies. More recently, there has been a tendency to reconsider how to focus on form in communicative syllabuses.

The emergence of alternative approaches to syllabus design led to a review of how second languages should be taught. The methodology most often associated with structural syllabuses—the audiolingual method—is underpinned by a behaviorist view of learning. Typically, this involves the presentation of paradigm grammatical structures, followed by repetition and drill with a view to the production of error-free sentences: practice makes perfect. The written language is used largely to reinforce or consolidate language learned orally. In CLT, by contrast, the emphasis is on learning by doing—on communicating fluently rather than accurately, and appropriately in terms of the social context in which the language is used. The four language skills of speaking, listening, reading, and writing are interwoven in meaningful classroom activities. The topics, contexts, and interactional patterns adopted in CLT classrooms reflect the learners' backgrounds and learning needs. The roles of learner and teacher are correspondingly different. In audiolingualism, learners are reactive: they listen to the input and respond by repeating, manipulating, transforming, and memorizing the linguistic data under the direction of the teacher. They are not encouraged to initiate interactions, because this

may lead to error. By contrast, learners in a typical CLT classroom are expected to be mentally proactive in forming and testing their own hypotheses about language and socially active in using the language with fellow learners. The audiolingual teacher models the language input, controls the direction and pace of practice, judges the accuracy of the learners' language behavior, and corrects errors as they occur. The CLT teacher is expected to participate in, rather than dominate, the process of learning—and to exemplify a model communicator, to determine and respond to the culture and needs of the learners, to facilitate group learning (for example, by identifying and enhancing the students' learning styles and strategies), and to organize a variety of resources, among which interactive computer-assisted language learning is becoming increasingly salient.

The emphasis on communication has led to a shift in how to evaluate second-language learning. Knowledge of the discrete phonological, syntactical, and lexical contents of a structural syllabus may be reliably measured by test formats such as multiple-choice questions, gap-filling tasks, sentence completion, and structural transformation exercises. However, the validity of such formats to assess communicative ability is questioned. Rather than measuring knowledge in terms of pass/fail marks, the focus is now on providing descriptive profiles on the basis of criteria-referenced performance in authentic oral and written interaction. Also, based on the assumption that candidates may operate at differential levels of productive or receptive competence, there is now a tendency to evaluate language skills separately, sometimes by portfolios of work and computer-adaptive assessment rather than pencil-and-paper tests. The contexts, materials, and tasks used to evaluate oral or written communication are increasingly designed to reflect the specific purposes of the learners.

References

Brown, H. Douglas. 1980. Principles of language learning and teaching. Englewood Cliffs, NJ: Prentice-Hall Regents; 4th edition, 2000.

Cohen, Andrew D. 1994. Assessing language ability in the classroom, Boston, MA: Heinle & Heinle.

———. 1998. Strategies in learning and using a second language. London and New York: Longman.

Harmer, Jeremy. 1998. How to teach English: an introduction to the practice of English language teaching. Harlow: Longman.

Hinkel, Eli (ed.) 1998. Culture in second language teaching and learning. Cambridge: Cambridge University Press.

Levy, Michael. 1997. Computer-assisted language learning: context and conceptualization. Oxford: Clarendon Press.

Odlin, Terence (ed.) 1994. Perspectives on pedagogical grammar. Cambridge: Cambridge University Press.

Paul, Peter. 1993. Linguistics for language teaching. Melbourne: Macmillan.

Richards, Jack C., and Theodore R. Rodgers. 2001. Approaches and methods in language teaching, 2nd edition. Cambridge: Cambridge University Press.

Richards, Jack C., and Richard Schmidt. 2002. Longman dictionary of language teaching and applied linguistics, 3rd edition. Harlow: Pearson Education Ltd.

Tomlinson, Brian (ed.) 1998. Materials development in language teaching. Cambridge: Cambridge University Press.

van Lier, Leo. 1996. Interaction in the language curriculum. London and New York: Longman.

ROGER BARNARD

Semantic and Discourse Typology

Semantic typology deals with the different types of semantic categories that can be found in natural languages, whereas discourse typology deals with different types of discourse. A very appealing approach to the classification of semantic categories that can be found in the lexicon is Anna Wierzbicka's proposal, which is based on the assumption of the existence of a closed set of semantic primitives that can be used for defining any concept in any language. According to this proposal, the meaning of a word does not depend on the meaning of other words, but rather on a configuration of the semantic primitives. The list she proposes is the following:

Substantives: *I, you, someone/person, something/thing, body*
Determiners: *this, the same, other*
Quantifiers: *one, two, some, all, many/much*
Evaluators: *good, bad*
Descriptors: *big, small*
Mental predicates: *think, know, want, feel, see, hear*
Speech: *say, word, true*
Actions, events, motion: *do, happen, move, touch*
Existence and possession: *there is, have*
Life and death: *live, die*
Time: *when/time, now, before, after, a long time, a short time, for some time, moment*
Space: *where/place, here, above, below, far, near, side, inside*
Logical concepts: *not, maybe, can, because, if*
Intensifier, augmentor: *very, more*
Taxonomy, partonomy: *kind of, part of*
Similarity: *like*

Other classifications are based not on the existence of semantic primitives that can be combined to form definitions, but rather on the existence of semantic fields that group the different concepts that can be expressed through the vocabulary of a language.

The *Longman lexicon of contemporary English* uses the following semantic fields, which are further subdivided:

Life and living things
The body: its functions and welfare
People and the family
Buildings, houses, the home, clothes, belongings, and personal care
Food, drink, and farming
Feelings, emotions, attitudes, and sensations
Thought and communication, language and grammar
Substances, materials, objects, and equipment
Arts and crafts, science and technology, industry and education
Numbers, measurement, money, and commerce
Entertainment, sports, and games
Space and time
Movement, location, travel, and transport
General and abstract terms

These and other groupings can be proposed for content vocabulary, in contrast to grammar words, which are used only for expressing grammatical relations within the sentence. Another approach to a semantic classification takes into account the semantics of the different elements that make up the sentence. It is assumed that in any sentence, there is a predicate (the verb) and several arguments that take up semantic roles. These semantic roles are verb specific (giver, thinker, hearer, etc.), but they can be grouped into more general relations like Agent (entity that performs an action), Experiencer (entity that experiences something, as a 'cognizer', 'perceiver', or 'emoter'), Recipient (entity that receives something), Stimulus (entity that is the object of experience), Theme (entity that is located or that undergoes a change of location), and Patient (entity that undergoes a process or an action initiated by another entity). These semantic

roles can be reduced to only two: Actor and Undergoer. Sentential semantic roles can be found in any language; they only differ in the way in which they are expressed in the grammar of every particular language. There exist many classifications of semantic roles, and each grammatical theory has its own list, so that we may find other role names, like Source, Path, Goal, Location, Object, etc.

In contrast to semantic typologies, discourse typologies are more concerned with broader units, related to discourse. 'Discourse' is a term that covers many kinds of communication. According to the possibilities available, we can establish different general types. A first distinction can be made between oral and written discourse. Oral discourse is characterized by the direct contact made by the participants in the interaction and its immediacy. Oral communication takes place 'in real time', that is, the message is decoded by the receiver as soon as it is encoded into the oral message by the speaker. In such a communicative situation, the message is also conveyed with the help of means other than language, such as hand gestures, intonation, posture, etc. In contrast, written discourse is not so immediate and there is no simultaneous participation of the writer and the reader in the communicative situation. It may take a very long time since the writer writes the message until the reader reads it. Everything should be encoded into the text and may be subject to a much closer scrutiny than the oral message, which means that it requires a more careful production. Another distinction that should be taken into account is the one between everyday and literary language. It is obvious that literary language involves a more careful discourse, which affects organization and the use of vocabulary and discourse markers. However, this is still a basic kind of grouping, which would be previous to a more detailed classification.

Closer to the semantic typologies presented are some classifications that have been proposed for analyzing the internal structure of discourse, such as a classification of different semantic relations within discourse. Discourse fragments can be related according to temporal relations (chronological sequence and temporal overlap), matching relations (contrast and comparison), cause–effect relations, truth-and-validity relations (statement-affirmation, statement-denial, denial-correction, concession-contraexpectation, etc.), alternation relations (contrastive and supplementary), bonding relations (coupling, exemplification, exception), paraphrase, amplification (specification, exemplification), setting relations, etc.

As for the external characterization of discourse, there have been many different attempts to establish an exhaustive classification of discourse types, but all of them accomplish this task partially. A good starting point for a discourse classification would be to take into account the different functions that any discourse may have, which depend on the objectives. The objective of discourse may be information (which is the basis of informative discourse), expression (the basis of narrative discourse), and persuasion (the basis of argumentative discourse). All three types normally appear in combination, but they serve to give some foundation to a more sophisticated classification of discourse types. One of those classifications is Werlich's, in which there are five basic forms underlying all discourse types: descriptive, narrative, explanatory, argumentative, and instructive. These basic forms are related to innate categorization possibilities that human beings have, and they are characteristically expressed by different types of sentences.

These five basic forms have two different methods of presentation: subjective, which shows the writer's perspective, and objective, which is subject to verification by the readers. This gives the following discourse types with subjective presentation: impressionistic description (descriptive), report (narrative), essay (explanatory), comment (argumentative), and instructions (instructive). If the method of presentation is objective, then we have the following types: technical description (descriptive), news story (narrative), explication (explanatory), argumentation (argumentative), and directions, rules, regulations, and statutes (instructive). All of these types have characteristic sentence structures. A further subdivision of discourse types can be carried out according to the 'channel' involved (oral channel vs. written channel).

Another type of classification is based on the discourse situation. Steger et al. distinguish six discourse types according to six types of discourse situation: presentation, message, report, public debate, conversation, and interview. These types are characterized by certain features present in the discourse situation. For some of them (presentation, message, report) there is only one speaker, whereas for others (public debate, conversation, interview) there are multiple speakers. The rank of the actors involved can be equal (in public debate and conversation) or unequal (in the rest). The theme can be predetermined (in all except conversation) or not predetermined (in conversation). Finally, the method of theme treatment can be descriptive (in message and report), argumentative (in presentation, public debate, and interview), or associative (in conversation).

References

Chafe, Wallace L. 1982. Integration and involvement in speaking, writing and oral literature. Spoken and written language. Exploring orality and literacy, ed. by Deborah Tannen. Norwood, NJ: Ablex.

Crombie, Winifred. 1985. Process and relation in discourse and language learning. Oxford: Oxford University Press.

Goddard, Cliff. 2001. Lexico-semantic universals: a critical overview. Linguistic Typology 5–1.

Renkema, Jan. 1993. Discourse studies, Amsterdam and Philadelphia: John Benjamins.

Steger, Hugo, et al. 1974. Redekonstellation, Redekonstellationstyp, Textexemplar, Textsorte im Rahmen eines Sprachverhaltensmodells. Begründung einer Forschungshypothese. Jahrbuch Gesprochene Sprache, ed. by H. Moser. Düsseldorf: Schwann.

Van Valin, Robert D. 2001. An introduction to syntax, [Chapter 2, Grammatical relations.] Cambridge: Cambridge University Press.

Werlich, Egon. 1982. A text grammar of English. Heidelberg: Quelle and Meyer.

Wierzbicka, Anna. 1996. Semantics: primes and universals. Oxford: Oxford University Press.

CARLOS INCHAURRALDE

Semantics

Semantics is the study of the meaning of signs and representations, both mental and linguistic. The ultimate target of semantics is the construction of a general theory of meaning.

Following a traditional taxonomy, we can distinguish three different kinds of signs depending on the nature of the relationship between the sign form and its meaning: indexical, iconic, and symbolic signs. Indexical signs are those in which the sign or representation is causally connected to its meaning. The redness of a face, for instance, is an indexical sign of anger or irritation. Certain skin spots are indexical signs of an illness such as measles. This connection is sometimes labeled, following Grice (1957), the natural meaning of a sign. In the case of icons, the relationship between the sign and its meaning is based on a perceptual similarity. The map of a geographical area is thus an iconic sign of the territory it represents. Symbols, finally, are signs whose relationship with their meaning is established through a conventional rule. A red light symbolically, i.e. conventionally, represents the obligation to stop. Individual expressions and sentences of natural languages are paradigmatic examples of this semantic category. Thus, the sentence 'snow is white' is the symbol we use in English to represent the fact that snow is white.

The appeal to this taxonomy implicitly involves the idea that the different relationships that connect the meaning of a sign to its form should be explained following different theoretical principles. Thus, the semantics of indexical signs would depend crucially on a theory of causal relationships; the semantics of icons would depend on the similarity principles that determine perceptual structures; and the semantics of symbols would be characterized in terms of the principles that regulate and constrain the conventional rules that help individuate the symbols' meaning. However, these demarcations are not as clear as they seem. This is especially the case in contemporary semantic theories, which often invoke causal (e.g. Fodor 1990) and perceptual factors in order to explain the meaning of symbols —and not just indexical and iconic signs.

From a philosophical point of view, symbols constitute the fundamental category within a semantic theory. This is the case, at least, within the so-called contemporary analytic tradition in philosophy. If we specify the range of phenomena that a semantic theory addresses following the scheme 'A interprets B as representing C', where 'A' is the interpreter, 'B' is an object, event, symbol or representation, and 'C' is the meaning of B, we can establish some important theoretical demarcations. The formal structure of B is studied by syntax. The relationship between B and its interpreter(s) is the domain of pragmatics. A semantic analysis would shed light on the nature and structure of C as well as on the nature and structure of the relationship between B and C.

Despite these basic differences, there are multiple and complex relationships between these three different disciplines. Thus, the array of meanings that C can represent is constrained by the syntax of B within a particular linguistic or representational system. In order to represent an event, for instance, the type of symbol should be a sentence, not a name. In this sense, logic, an essentially syntactic discipline, is sometimes considered a part of semantics because it aims at explaining the structural properties of the expressions of a language.

If the boundary between syntax and semantics can be fuzzy, the demarcation between semantics and pragmatics is even more problematic. The most radical conception in this sense, namely a conception according to which the notion of meaning of a symbol can be reduced to the conditions of its use (Wittgenstein 1953), denies that there is a distinction at all.

These difficulties notwithstanding, there are indeed certain questions that are the proper domain of semantics. Among these questions, the most prominent is perhaps that of clarifying and determining the very nature of the notion of meaning. In principle, the meaning of an expression could be thought as just what the expression stands for, i.e. its reference. This is a technical sense of reference understood as a function that assigns, to the symbols of a language, different extralinguistic entities taken from a particular domain of interpretation. This definition coincides with the characterization of reference as a semantic value. Its only peculiarity consists in including—as an additional layer of complexity—the idea of a domain of interpretation. In the case of natural languages—i.e. languages such as English or Spanish—this interpretation is already given and the task of a semantic theory is to specify the interpretation already imposed by the way in which the language is used. However, in the case of nonnatural or artificial languages (e.g. the language of logic or mathematics), we have to fix a nonempty domain of objects, if we want the symbols of those languages to mean anything at all, i.e. if we want to interpret those languages or—as is sometimes said when talking about formal languages—if we want to have a model of those languages. Thus, this characterization of the notion of reference responds to the basic requirements of what is known, in mathematics, as model theory (see e.g. Hodges 1993). Although this, in itself, is no more than a branch of mathematics that studies the relationships between different types of languages and different types of mathematical structures, it has strong implications and applications in the territories of semantics, linguistics, and the philosophy of language.

The notion of reference, however, does not completely exhaust the meaning of a symbol. Frege (1892) introduces an additional and even more fundamental semantic concept to complement that of reference: the concept of sense. The need to invoke this new semantic dimension when trying to determine the meaning of a symbol becomes clear whenever we address the issue of explaining why true identity statements involving different proper names (e.g. 'Hesperus is Phosphorus') which refer to the same object (Venus) differ from identity statements that involve the same name (e.g. 'Hesperus is Hesperus'). If the meaning of a proper name (e.g. 'Hesperus') were just its reference (i.e. the planet Venus), the two kinds of identity statements should have the same cognitive value. Since this is not the case—since we might learn something we did not know when we learn that Hesperus is Phosphorus—we have to admit that the meaning of a proper name cannot be just its reference.

The same two concepts—sense and reference— have been used to analyze not just proper names but also other kinds of expressions, such as declarative sentences. In this case, the meaning (sense) of a sentence is characterized as its truth conditions, i.e. as what the world would be like if the sentence were true. The reference of the sentence is its truth value. There are only two truth values in classical semantics: truth and falsity, and establishing the truth value of a sentence, unlike determining its meaning, is an empirical matter. The idea of the meaning of a sentence being its truth conditions has to be amended to cover nondeclarative sentences such as imperatives or questions, but it has been a central notion in all contemporary semantic theories. Thus, Donald Davidson's primary insight, and his major contribution to semantic theory, was to adapt a theory of truth developed by the mathematician Alfred Tarski (1944) to the problem of developing a theory of meaning for natural languages. Tarski assumed the idea of meaning—the idea that a sentence could be the translation of another sentence—to get at the notion of truth. Davidson inverts Tarski's scheme and uses the idea of truth as basic in order to develop a characterization of meaning in terms of truth conditions (Davidson 1984).

Such a characterization of meaning also helps establish what inferences can be legitimately drawn from each sentence, since a valid inference or argument is one in which the truth of the conclusion follows from the truth of the premises. The notion of truth—the core notion in logic—and the notion of meaning—the core notion in semantics—are thus intimately related. In fact, mathematical logic has provided the basic theoretical tools for the development of some of the most interesting semantic theories of this century. In particular, the development of first-order predicate logic by Gottlob Frege opened the door for many applications of semantics within areas such as computer science and artificial intelligence.

Despite this initial success, contemporary semantic theorists tend to question the plausibility of applying the model of first-order logic to natural languages. Natural languages have a much more complicated syntax than the language of first-order logic: expressions are often ambiguous, and there are many kinds of sentences other than declarative sentences. A first attempt at providing an alternative account was developed by Montague (1974), who advocated the use of a more powerful logic (intensional logic) as a way of solving some of the problems encountered within the original Fregean treatment. Montague's proposal allows us to characterize the meaning of sentences involving the idea of necessity or possibility and also of those sentences that express propositional attitudes such as belief or desire. However, even Montague's treatment remains heavily embedded in the formalist approach set up by the mathematical view of semantics and it too has been called into question by more recent semantic theories.

The characterization of the notion of sense, especially when applied to common names, i.e. expressions such as gold, tiger, etc., has also been harshly criticized over the last few decades. This notion was traditionally characterized in terms of a set of necessary and sufficient conditions that determine the reference of an expression. The basic idea of the recent critics is that we can ascribe to the average speaker of a language knowledge of the meaning of an expression without assuming thereby that the speaker knows the set of necessary and sufficient conditions that determine the reference of such an expression. The alternative proposal is to substitute the notion of stereotype for the idea of sufficient and necessary conditions. A stereotype is a mental representation of a paradigmatic instance of an object or property with which the subject has had some kind of experience (Putnam 1975).

In contemporary semantics and philosophy of language, the notions of narrow and wide content have come to substitute those of sense and reference. Although intimately related to the original ones, the new notions have a psychological slant which was not present in earlier versions of the same idea, and are basically used in the context of specifying the meaning of a subject's mental representations.

On the one hand, the narrow content (or meaning) of a mental state is constituted by the intrinsic (i.e. context-independent) properties of the individual who is in that mental state. Internalism is the point of view characterized by the thesis that such contents are individuated without essential reference to the subject's physical and social environment. On the other hand, the externalist claims that contents are individuated 'widely', i.e. by reference to the subject's environmental or social context.

To begin with, we said that the aim of semantics is the development of a theory of meaning. It is important to stress now that one of the most important topics in this discipline is the study of the empirical evidence that verifies—or falsifies—such a theory. This empirical dimension brings semantics into the domain of naturalism, i.e. the domain of empirical science. Dominant naturalistic theories depict semantic properties as properties of some other (more primitive, less problematic) kind since semantic properties are not commonly held to be part of the basic ontological furniture of the world. Most theorists engaged in naturalization projects for semantics thus assume that the program of naturalization demands a higher-to-lower, top-to-bottom, kind of explanatory strategy. They believe, in other words, that the nonsemantic properties on which semantic properties depend, belong to what are intuitively lower levels of description than the semantic level itself (contenders include biological properties; see Millikan 1984). The reductionist flavor is unmistakable here. Achieving a clear understanding of 'naturalism' as applied to semantics and deciding whether a naturalistic semantics is possible are probably two of the most important issues for contemporary philosophy of language and philosophy of mind.

References

Davidson, Donald. 1984. Inquiries into truth and interpretation. Oxford: Clarendon Press.

Fodor, Jerry. 1990. A theory of content and other essays. Cambridge, MA: MIT Press.

Frege, Gottlob. 1982. Über Sinn und Bedeutung. Zeitschrift für Philosophie und philosophische Kritik 100; Also in Translations from the philosophical writings of Gottlob Frege, ed. by Peter Geach and Max Black. Oxford: Blackwell, 1960.

Grice, Paul. 1957. Meaning. Philosophical Review 64.

Hodges, Wilfrid. 1993. Model theory. Cambridge: Cambridge University Press.

Millikan, Ruth. 1984. Language, thought, and other biological categories: new foundations for realism. Cambridge, MA: MIT Press.

Montague, Richard. 1974. Formal philosophy: selected papers of R. E. Montague, ed. by R.H. Thomason. New Haven: Yale University Press.

Putnam, Hilary. 1975. The meaning of meaning. Minnesota studies in the philosophy of science, Vol. VII: language, mind and knowledge, ed. by Keith Gunderson, Minnesota: University of Minnesota Press.

Tarski, Alfred. 1944. The semantic conception of truth and the foundations of semantics. Philosophy and Phenomenological Research 4.

Wittgenstein, Ludwig. 1953. Philosophical investigations, transl. by Elisabeth Anscombe, Oxford: Blackwell.

JOSEFA TORIBIO

Semiotics

Semiotics is the science of signs. Even though the study of signs has been a constant in the history of thought, we owe the rise of semiotics as a discipline in the modern world to the ideas of the Swiss linguist Ferdinand de Saussure and the American philosophers Charles S. Peirce and Charles W. Morris.

In the tradition of Saussure, a sign is composed of the signifier, i.e. the form that the sign takes, and the signified, i.e. the concept or object it represents. The signifier is characterized as the physical form of the sign. It is important to keep in mind the distinction between the sign *vehicle*, which is just the form in which the sign appears (spoken or written form, for instance), and the sign itself, which includes both the signifier and the signified. The signified has, within this tradition, not a physical but a psychological status. It is interpreted as some kind of mental construct, the psychological impression that the signifier makes on the interpreter of the sign. Given this characterization of how signs represent, the idea of signs referring to objects in the world is excluded from this particular view of semiotics. Signs do refer, but only to mental concepts and psychological entities constructed by the interpreter.

This mentalist understanding of the notion of reference of a sign (the signified) seems to support an idealist philosophical interpretation of reality, i.e. an interpretation according to which there are no ultimate real objects and events in the world regardless of how we construct them throughout the use of signs. It is important to notice, though, that this rather psychologized view of the referent or signified of a sign is not shared by all semioticians. Charles Peirce's proposal, for instance, views the referent of a sign as something with its own ontological status, something external to the sign's interpreter. Peirce's model, by characterizing the referents of signs as something external to the interpreter, as something with an independent ontological weight, seems to suggest a more realist philosophical interpretation of the world according to which reality is composed of objects and properties which stand on their own and independently of our means of describing them or of knowing them via a system of signs.

Saussure calls the relationship between the signifier and the signified the signification. For Saussure, the relationships between signifiers and their signifieds are ontologically arbitrary, but once these relationships have been established by the use of a community throughout history, such arbitrary connections cannot be changed. But regarding this issue of arbitrariness, not all semioticians agree. Most would nonetheless allow that there are degrees of arbitrariness in how the signifier and the signified relate to each other. Thus, following Charles Peirce's model, we can establish three different kinds of signs depending on how the signs vehicles relate to their referents. From the most to the least conventionally based relationships, we find symbols, icons, and indexes. Symbols are purely conventional, i.e. there is no perceptual or naturally recognizable relationship between the signifier and the

signified. An iconic sign is one in which the signifier physically or perceptually resembles the signified (e.g. a portrait or a sculpture). Indexical signs are those in which the signifier is physically or causally connected to the signified, such as smoke to fire, facial spots to measles, etc.

When we talk about signs in this context, we are thus referring not just to linguistic symbols. A sign is anything that stands for something else. Thus, semiotics can be characterized as a theoretical approach to such diverse phenomena as images, sounds, paintings, rituals, gestures, etc., as well as languages. The presence of the theoretical apparatus of linguistics is, however, quite significant since linguistic symbols are still considered paradigmatic within this discipline. In this sense, Claude Lévi-Strauss, a contemporary anthropologist who sees anthropology as a branch of semiotics, claims that 'language is the semiotic system par excellence; it cannot but signify, and exists only through signification' (Lévi-Strauss 1972: 48). Semiotics enjoyed its golden era within the philosophical movement known as Logical Positivism in the 1930s. C.W. Morris' *Foundations of the theory of signs* probably provides the most complete characterization of this theory, presenting it as a science whose main aim is to provide the foundations for any other particular science about signs.

Having such a wide range of phenomena as its subject makes it hard to establish a precise demarcation between semiotics and the other disciplines with which it shares some of the same objects of study, e.g. disciplines such as linguistics, logic, mathematics, esthetics, and anthropology—all of which address, in one way or another, the general phenomenon of signs. Nor is it obvious that the proper theoretical status of semiotics is that of a science, i.e. that we should consider semiotics as having the same theoretical status as other empirical sciences, such as biology or chemistry. Semiotics sometimes appears as a style of philosophical investigation, sometimes as a theory, and some other times as a methodology. In order to gain a better understanding of this discipline, we shall focus here on two issues: the domain of semiotics and its theoretical status.

Semiotic processes or events involve, according to Charles S. Peirce, three factors: a sign, an object, and an interpreter. Depending on which factor gets isolated for study, semiotics splits into three different branches: pure grammar (the study of signs themselves), logic (the study of the object of the sign), and rhetoric (the study of the sign's interpreter). But what characterizes semiotics and makes it different from the other disciplines is that the interrelation between these three factors is absolutely essential for the individuation of semiotic phenomena. Charles W. Morris (1938, 1946)

likewise stresses the dynamic interrelation between sign, denotation, and interpreter. From Morris' point of view, no object, event, or particular action can be considered, in itself, a sign, a denotation, or an interpreter. The properties that characterize each of these entities are relational properties. Such properties are acquired only through a semiotic process, i.e. a process through which something becomes a sign for someone (cf. Morris 1946: 353). Morris establishes four components in this process: the sign (vehicle), the denotation of the sign, the organism for which something is a sign (the interpreter), and the effect the sign produces in the interpreter. Among the multiple examples that Morris provides to illustrate the role played by each of those factors, perhaps the clearest is that of a traveler who, ready to go to a specific geographical area, receives the letter of a friend describing that region. The traveler then packs his luggage in accordance with the information provided by his friend in the letter. In that example, the letter is the sign, the geographic conditions are the denotation of the sign, the traveler's preparations regarding what to pack are the effect that the sign produces in the interpreter, and the traveler, of course, is the interpreter. The most important idea is that none of those factors would belong to these categories were it not for the interrelationships among them (cf. Morris 1938: 3–4).

Despite this theoretical interdependence, Morris, as well as Peirce, concedes that we can isolate the relationships between such categories and define different branches within semiotics. Thus, if we are mainly interested in the relationships between the signs themselves, we focus upon the syntactic dimension of semiotics. We focus on the semantic dimension if we are mainly interested in the relationships between the signs and their denotations and upon the pragmatic dimension if our focus is upon the relationships between signs and their interpreters (Morris 1938: 6–7). These partitions can, in principle, help establish clearer theoretical connections between semiotics and other disciplines. Thus, linguistics and logic are about the syntactic semiotic dimension. The object of semantics is the relationship between signs and their objects—the semantic dimension of semiotics. Pragmatics, finally, is the most problematic discipline whenever we try to assign to it a specific semiotic dimension. The reason for this is that the study of the relationships between signs and their interpreter(s) covers a complex field, which includes psychological, biological, and sociological aspects of sign manipulation. Psychoanalysis, the sociology of knowledge, and even some approaches within esthetics or anthropology all touch upon this pragmatic dimension of semiotics. Furthermore, the problem with this taxonomy is that it seems to suggest that whenever we are doing syntax or semantics, we can ignore the role of the interpreter. However, such a suggestion is highly controversial, especially if we realize that the complex notion of the meaning of a symbol seems to include a variety of aspects related to the interpreter's use of those symbols, thereby blurring the boundaries between semantics and pragmatics. Morris himself downplayed this taxonomy in his later work *Signs, language and behavior*, where he introduces it just as a general suggestion and only in the last chapter. The problem of distinguishing pragmatics and semantics has been and still is one of the central problems in linguistics and the philosophy of language.

So far, we have been talking about the object of semiotics. But what can we say about its theoretical status? For some authors, semiotics ought to be placed at the same level of philosophy of language (where 'language' includes nonverbal languages). Umberto Eco (1984) is perhaps the most fervent advocate of this view. For him, the philosophical character of semiotics turns it into a discipline without any predictive power, and in which the object of study is also always mediated by the general presuppositions of the theory itself (Eco 1984: 10–3). However, this is not a widely shared view. Especially when we talk about specific semiotics (e.g. narrative semiotics, folk semiotics, animal behavior, and human institution semiotics, etc.), the descriptive slant of the discipline is stressed and, with it, its empirical character. Finally—and without considering these distinctions exhaustive—semiotics is sometimes considered more as a general methodology or as the sum of different semiotic methods. From this point of view, the notion of sign is applied not only to linguistic objects but also to objects such as buildings, rituals, landscapes, and star formations. Among the semiotic methods, we can distinguish at least three: formalization, linguistic analysis, and method of interpretation.

The method of interpretation—especially the formalization of natural languages—stresses the syntactic properties of signs and what is sometimes called 'operational meaning': 'operational' because it is normally characterized in terms of rules that specify the admissible forms for manipulating signs. Once signs are stripped of their semantic and pragmatic properties, they constitute a purely syntactic skeleton, ready to be dressed with different interpretations.

The method of linguistic analysis consists in the reconstruction of the deep structure of the analyzed expression and in the enumeration of all their semantic and pragmatic properties. The use of this method is thus complementary to the method of formalization and it allows us to find out what an expression or a particular piece of text really means and what information can be derived from its meaning. The use of this method lies at the heart of so-called 'structural

linguistics' (Saussure 1916). It has also been widely applied to the sciences, paying special attention to the process of elaboration of particular scientific languages and to the creation of new terms.

The method of interpretation consists fundamentally in the symbolic treatment of objects, events, or phenomena that are not themselves signs (e.g. food, rituals, human faces, etc.). When we interpret them in this way, we direct our attention beyond the object or event itself. In so doing, objects, events, and phenomena that are not in principle symbolic at all become an instrumental part of our system of communication. This is probably the most familiar version of what is often understood by 'semiotics', namely, a treatment that can be considered an extension of hermeneutics in the sense of being applied to phenomena that are not intrinsically symbolic. This method is highly relevant for disciplines such as psychology or sociology since one characteristic type of explanation within these sciences consists in bringing to light intentions and motives that are only indirectly expressed in human behavior and human intellectual creations.

References

Bouissac, Paul (ed.) 1998. Encyclopedia of semiotics. Oxford: Oxford University Press.

Eco, Umberto. 1984. Semiotics and the philosophy of language. London: Macmillan Press.

Jackobson, Roman. 1980. The framework of language. Ann Arbor, MI: Slavic Pub.

Lévi-Strauss, Claude. 1972. Structural anthropology. Harmondsworth: Penguin.

Morris, Charles W. 1938. Foundations of the theory of signs. Chicago: University of Chicago Press.

Morris, Charles W. 1946. Signs, language, and behavior. New York: Prentice-Hall.

Nöth, Winfried. 1990. Handbook of semiotics. Bloomington: Indiana University Press.

Peirce, Charles S. 1966. Selected writings, New York: Dover.

Saussure, Ferdinand de. 1916. Cours de linguistique générale (Course in General Linguistics). Paris: Payot and London: Fontana/Collins, 1974.

Sebeok, Thomas (ed.) 1977. A perfusion of signs, Bloomington: Indiana University Press.

JOSEFA TORIBIO

See also **Levi-Strauss, Claude; Peirce, Charles; Saussure, Ferdinand de**

Semitic Languages

Four of the five literary Semitic languages that underlie great civilizations of Asia, Africa, and Europe—Arabic, Aramaic, Geʻez or Ethiopic, and Hebrew—are covered elsewhere, so here we concentrate on the fifth, Akkadian; on the less well-known Semitic languages of antiquity; and on what they have in common and how they relate to each other. Semitic itself is one of five or six branches of the Afroasiatic phylum, the only one spoken principally in (southwest) Asia.

General Characteristics

For a thousand years or more, linguists (beginning with the Arab grammarians) have taught that words in Semitic languages are formed from 'roots' consisting of three consonants and no vowels at all, and only 'modifications' of the 'root meaning' are indicated by vowel 'patterns'. But this seems to be an artifact of the writing systems used by the Arab (and Hebrew and Syriac) grammarians, which notated only the consonants, with the short vowels included only optionally and with incidental-looking marks. Moreover, it relates only to verbs, and not to 'primary nouns'. In 1934, an Indo-Europeanist, Louis H. Gray, essayed a description of Semitic that eschewed the root-and-pattern model and observed verb roots of the form CCVC—like many other languages—rather than the unique CCC. (See Table 1. This view has recently been adopted independently by several scholars.)

Typical of the Semitic languages is their way of expressing a relationship between two nouns. Rather than marking the possessor, as in *the king's peace*, Semitic puts the possessed into the CONSTRUCT STATE, as in Akkadian *šulum šarri*, literally 'peace-of king'; cf. *šulmu* 'peace', *šarru* 'king'. These words also illustrate two of the three noun cases: nominative ending in *-u*, genitive (possessive or adnominal) ending in *-i*, and accusative (objective or adverbial) *-a*. Nouns in Semitic have two genders: masculine or feminine.

The morphological typology of Semitic is inflecting; the syntactic typology is basically Verb–Subject–Object, even in those languages where the verb usually comes at the end of a clause: Ethiopic, presumably under the influence of Cushitic languages, and Akkadian, under the influence of Sumerian.

TABLE 1 Consonant Correspondences, Prefix Conjugation, and Verb Stems[1]

Translation	Akkadian	Hebrew	Syriac	Geʿez	Arabic	Proto-Semitic
I guard	*aṣṣuru*	*ʾeṣṣor*	*ettur*	**ʾinṣir*	*ʾanzuru*	*ʾa-nθur*
We grind	*niṭên*	*niṭhan*	*neṭhan*	*niṭhan*	*naṭhanu*	*na-ṭhan*
You (m.sg.) are lame	*tēgir*	*tahăgor*	*tehgar*	*cf. tɨhgɨl*	*tahjuru*	*ta-ḫgur*
You (m.pl.) harvest	*tēṣidā*	*(maʿăṣad* 'axe')	*tehṣdūn*	*tiʿdɨdū*	*taḥḍidūna*	*ta-ʿ/ḥḍid-ū*
You (f.sg.) seize	*tāḫuzī*	*tōḥăzī*	*te(')ḥzīn*	*tiʾḥazī*	*taʾḫuðīna*	*ta-ʾḫuð-ī*
You (f.pl.) enter/set	*tērubā*	*(ʿereb* 'evening')	*teʿrbān*	*tɨʿrabā*	*tagrubna*	*ta-ǵrub-na*
He binds	*īsir*	*yeʾĕsor*	*ne(')sor*	*yɨʾsɨr*	*yaʾsiru*	*ya-ʾsir*
They (m.) go	*illikū*	*yahalkū*	**nehlkūn*	**yɨhlikū*	*yahlikūna*	*ya-hlik-ū*
She places	*tašīm*	*tašīmī*	*tsīm*	*tɨšīm*	*tašīmu*	*ta-śyim*
They (f.) hunt	*tiṣūdā*	*taṣōdnā*	*tṣūdūn*	**tɨṣūdā*	*taṣūdūna*	*ta-ṣwud-na*
To sing (D)	*zummuru*	*zammēr*	*mzammārū*	*zammɨrō*	*tazmīrun*	*zmi/ur*
To cause to hear (Š)	*šušmû*	*hašmīăʿ*	*mašmāʿū*	*asmɨʿō*	*ʾismāʿun*	*šmaʿ*
To be twisted (N)	*naptulu*	*hippātēl*	*— (√ptl)*	*— (√ftl)*	*infitālun*	*ptil*
To be sought for (Gt)	*pitqudu*	*— (√pqd)*	*metpqādū*	*tafaqɨdō*	*iftiqādun*	*pqid*
To find one's match (Dt)	*mutaššulu*	*hitmaššēl*	*metmattālū*	*tamassɨlō*	*tamaθθulun*	*mθul*

[1]The correspondences yielding the 29 reconstructed Proto-Semitic consonants can be read in the rows, as can the prefixes. The last five rows show the most important derived stems: G(round) or Basic stem (in the top 10 rows), D(ouble)=Intensive, Š=Causative, N=Passive, t=Passive/Reciprocal. * marks roots not attested in this language or this stem. — (√…) marks stems not used in this language, with the consonant correspondences for the language.

Classification

The sole representative of East Semitic is Akkadian, which includes the dialects Assyrian, Babylonian, and the recently discovered Eblaite. West Semitic was understood to comprise Northwest (divided into Canaanite and Aramaic) and Southwest (the languages of South Arabia, both Ancient and Modern, and Ethiopia); Arabic used to be included with the Southwest group, but in recent years arguments were advanced that it belonged better with the Northwest group. The criteria for the classification are presented below, following Faber in Hetzron (1997).

Recently, however, *Stammbaum* or 'family tree' diagrams of language family relationships have fallen out of favor, as linguists have recognized the importance of social interaction in the shaping of languages, and a new model for the development of the Semitic languages has been offered (Kienast 2001). It suggests that the successive appearances of the languages in the historical, i.e. written, record represent successive attestations from a single, continually developing speech community (and after they make their entries, of course, each language

would develop independently). This approach helps explain how some characteristics (like noun case endings) can be found in the earliest (Akkadian) and latest (Arabic) forms of Semitic, but not in the intermediate forms (taken to be Ethiopic/South Arabian, Canaanite, and Aramaic in order).

Although Akkadian was deciphered in the late 1840s, its grammar was not well understood before the insights of the Czech-born Leipzig/Ankara/Chicago Assyriologist Benno Landsberger, which were codified first in a chapter by Gotthelf Bergsträsser (1928/1995) and then by Landsberger's student Wolfram von Soden (1952/1995); the earliest form of Akkadian received due attention only with the work of I.J. Gelb in the 1950s. Consequently, Akkadian was not well assimilated into the early comparative sketches and collections of data such as those by William Wright (1890), Theodor Nöldeke (1889/1911), and Carl Brockelmann (1908, 1908–1913) and even that by Landsberger's own teacher Heinrich Zimmern (1898). In these works—and in subsequent manuals, notably the widely used Moscati et al. (1964)—the reconstructed Proto-Semitic bears a very close relationship to Classical

Arabic. (Lipiński (1997) is not suitable for the beginner because, although detailed and up to date, it does not distinguish between the generally accepted opinion and the author's own; but its bibliography is extensive.) Only with the work of Gelb himself (1969) and of fellow Assyriologist Burkhart Kienast (2001) does a Proto-Semitic emerge that offers a more credible abstraction that might foreshadow East Semitic as well as West Semitic.

Even before the decipherment of Mesopotamian cuneiform had progressed very far, the language written with that script could be identified as Semitic (which in turn proved very useful in the decipherment). The clue was the identification of sets of words exhibiting perhaps the most distinctive trait of Semitic morphology: alternations of the type *iprus–iparras–paris*. The precise meanings of these verb paradigms turned out not to be exactly those familiar from Hebrew, Arabic, and Ge'ez, but the forms were identical.

In fact, the divergence in meaning of the verb forms is taken as the most salient characteristic for distinguishing West Semitic as a whole from Akkadian. All Semitic languages have just two basic verb forms: in Akkadian *iprus*, the 'preterite' or past, and *iparras*, the 'present' (persons are identified by prefixes; Table 1); *paris*, the 'stative', stands somewhat outside the system (persons are identified by suffixes; Table 2). But West Semitic has turned the equivalent of *paris*, e.g. Arabic *qatala*, into the 'perfect', denoting completed action (forms of *qatala* 'kill' are traditionally used to illustrate verb forms; this word is not found in Akkadian, and *parāsu* 'divide' serves instead). The forms with a lengthened middle consonant, e.g. *qattala*, are reserved for a quite different function, but one that (with a different selection of vowels) it already had in Akkadian.

This different function is another signal characteristic of Semitic: rich derivational morphology of the verb. English, for example, has pairs of semantically related words like *learn* and *teach*. But in Semitic, 'teach' is literally 'cause to learn': Akk. *lamādu*, *šulmudu*. Besides this CAUSATIVE, Semitic languages have an INTENSIVE—the derivation marked by the lengthened middle consonant—and one or two kinds of PASSIVE (and additional derivations appear in individual languages). Some examples are shown in Table 1.

The forms in Table 1 also exemplify the sound correspondences that both demonstrate the unity of the Semitic family and characterize its different branches. In East Semitic, the five gutturals merge into ' [ʔ], but the three most highly marked ones, *ǵḥ'* [ɣ ħ ʕ], color an adjacent *a* to *e*, while *h* and ' do not.

In Central Semitic, the EMPHATIC series (transcribed with the underdot) is realized with pharyngealization, rather than glottalization as in South and (probably) East Semitic. The nonpast *yaqattal* is replaced by *yaqtulu*. Within Central Semitic, Northwest Semitic is distinguished from Arabic by the change of almost every initial *w* to *y*. Within Northwest Semitic, Canaanite shows several common morphological innovations, and the change of *ā* to *o*. Aramaic displays a simplification and 'neatening' of the derived verb stem system, but its typical phonological change, the interdentals [θ ð] becoming stops [t d], is not found in the earliest attested texts. In Canaanite, the interdentals become sibilants [ʃ ʒ]. (Arabic preserves all the ancient distinctions except one of the sibilants. Semitists usually write *š ś ǵ ḥ* for [ʃ ɬ ɣ x].)

South Semitic's diagnostic morphological innovation is seen in Table 2. The *k* of the old first-person singular has spread to the second person (while in Northwest Semitic, the *t* of the second person has spread to the first). Southeast Semitic (Modern South Arabian) is not the direct descendant of Old (or Epigraphic) South Arabian; the latter is closer to Ethiopic Semitic than the former is to either of them.

TABLE 2 Person, Gender, and Number Suffixes on *paris/qatala* Forms

Person	Akkadian	Hebrew	Aramaic	Ge'ez	Arabic	Proto-Semitic
1 sg.	*-āku*	*-tī*	*-et*	*-kū*	*-tu*	*-ku*
1pl.	*-ānu*	*-nū*	*-n*	*-na*	*-nā*	*-na*
2m.sg.	*-ātu*	*-tā*	*-t*	*-ka*	*-ta*	*-ta*
2m.pl..	*-ātunu*	*-tem*	*-tūn*	*-kɨmmū*	*-tum*	*-tumu*
2f.sg.	*-āti*	*-t*	*-t*	*-kī*	*-ti*	*-ti*
2f.pl.	*-ātina*	*-ten*	*-tēn*	*-kɨn*	*-tunna*	*-tinna*
3m.sg.	—	—	—	*-a*	*-a*	—
3m.pl.	*-ū*	*-ū*	*-ū*	*-ū*	*-ū*	*-ū*
3f.sg.	*-at*	*-ā*	*-at*	*-at*	*-at*	*-at*
3f.pl.	*-ā*	*-ū*	*-ā*	*-ā*	*-na*	*-ā*

Specific Languages

Akkadian was spoken in Mesopotamia from before 2500 BCE until the early to mid–first millennium BCE and survived as a literary and then scholarly language into the early centuries CE (Geller 1997). (The latest known dated Akkadian text is from 75 CE.) Akkadian was written by impressing groups of wedges in soft clay shaped into 'tablets', which when they dried (or got baked when a building burned down) were quite permanent, and thousands were recovered when explorers began poking into the remains of ancient cities. Its decipherment in the late 1840s revealed the existence of an unsuspected Mesopotamian civilization of which bare hints were found in the Bible; early on, it was found to have a literature, a religion, and a mythology (including a Flood story) that illuminated not only the Bible but its own society. Even more, though, emerges from the tens of thousands of non-literary tablets recording ordinary transactions of the most varied kind. There are two scholarly dictionaries: Gelb et al. (1956–) and von Soden (1958–1981); Black et al. (2000) is a condensation of the latter.

The second half of the nineteenth century witnessed the discovery of a wide variety of inscriptions in Hebrew and other, closely related Canaanite languages, the first and foremost, and for a long time the earliest (c. 850 BCE), being the stela of King Mesha, in Moabite. Its script is very similar to that of Phoenician, which had been deciphered in the 1760s; unfortunately, the large number of inscriptions in Phoenician (c. 1000 BCE–500 CE), found scattered all round the Mediterranean, including its late variety, the Punic of Carthage and environs, are almost entirely brief and uninformative, so that little can be reconstructed of Phoenician political history. The sole monument in Moabite, however, is a historical inscription resembling contemporary accounts in Aramaic, Akkadian, and Hebrew. In recent decades, it has been joined by similar but less extensive materials in Ammonite and Edomite. All three languages are from east of the Jordan River. Akkadian tablets preserve references to thousands of people whose names are in Amorite, the West Semitic language of the famed King Hammurapi and his dynasty.

An epochal discovery in 1929 at Ras Shamra on the coast of Syria was clay tablets in what soon proved to be a Northwest Semitic language, Ugaritic. Its script is consonantal like Phoenician's, and not derived from Mesopotamian cuneiform despite the use of wedge-made letters. Dating from the fourteenth–thirteenth centuries BCE, Ugaritic is too early to display specifically Canaanite features, but again it preserved texts illuminating the religious background of the Bible. In 1975, farther east in Syria, a large number of clay tablets were found at Tell Mardikh, ancient Ebla. At first their Eblaite language, written in the familiar cuneiform of c. 2500 BCE, was interpreted as also representing a very early relative of Hebrew, but it was soon recognized as a form of Akkadian or as a separate East Semitic language used far to the west.

The South Arabian languages, Modern (Mehri, Ḥarsusi, Baṭhari, Hobyot, Jibbali, and Soqoṭri) and Old (Sabean, Qatabanian, Hadramauti, and Minean), were the first languages (after Phoenician) to be 'discovered' in modern times, in the 1830s. They occupy the southern fringes of the Arabian peninsula and nearby islands. All the proto-Semitic consonants are preserved, and the Old South Arabian script records the 29, and no vowels at all. The modern languages are not written; Mehri and Soqoṭri may have enough speakers not to be immediately endangered.

Semitic Writing

The scriptures of the three Western monotheistic religions are written in Semitic languages (Hebrew, Aramaic, and Arabic). But the Semitic languages hosted the development of the writing systems themselves that serve virtually every language outside East Asia. Mesopotamian cuneiform, after being adapted from its Sumerian original, was used for many languages throughout the ancient Near East (with the exception of Egyptian) over at least 2,000 years, first of all.

Cuneiform was cumbersome, though, requiring the memorization of several hundred characters during years of study. It was when the basic principle of Egyptian hieroglyphics—write only consonants—was applied in a highly simplified manner to an early West Semitic language that the sequence leading to the alphabets of the modern world got going. Little material survives from the earliest stages; by the time of Ugaritic, we see an adaptation of the letters to writing on clay. One variety is found to the south, used in Old North Arabic graffiti and Old South Arabian monumental inscriptions, and then in Ethiopic inscriptions and manuscripts.

The other variety emerges in Phoenician and Aramaic versions; Phoenician survives almost exclusively in its offshoot, the Greek alphabet, and its descendants including the Roman and Cyrillic. Scions of the Aramaic version serve Hebrew, Arabic, and Aramaic—and spread all across South, Southeast, and Inner Asia. The variety of ways in which vowel notation came to the originally consonantal scripts is a saga in itself.

References

Bergsträsser, Gotthelf. 1983. Introduction to the Semitic languages, transl. by Peter T. Daniels. Winona Lake, In: Eisenbrauns; revised, 1995 (German original, 1928).

Black, Jeremy, Andrew George, and Nicholas Postgate 2000. A concise dictionary of Akkadian. Wiesbaden: Harrassowitz.

Brockelmann, Carl. 1908. Kurzgefasste vergleichende Grammatik der semitischen Sprachen [Brief comparative grammar of the Semitic languages]. Berlin: Reuther & Reichard.

————. 1908–1913. Grundriss der vergleichenden Grammatik der semitischen Sprachen [Foundations of the comparative grammar of the Semitic languages], 2 vols. Berlin: Reuther & Reichard.

Gelb, I. J. 1952. Old Akkadian writing and grammar. Chicago: University of Chicago Press, 2nd edition, 1961.

————. 1969. Sequential reconstruction of Proto-Akkadian. Chicago: University of Chicago Press.

Gelb, I. J., et al. (eds.) 1956–. The Assyrian dictionary of the Oriental Institute of the University of Chicago, 21 vols. Chicago: Oriental Institute.

Geller, M. 1997. The last wedge. Zeitschrift für Assyriologie 87.

Gray, Louis H. 1934. Introduction to Semitic comparative linguistics. New York: Columbia University Press.

Hetzron, Robert (ed.) 1997. The Semitic languages. London: Routledge.

Kienast, Burkhart. 2001. Historische Semitische Sprachwissenschaft [Historical Semitic linguistics]. Wiesbaden: Harrassowitz.

Lipiński, Edward. 1997. Semitic languages: outline of a comparative grammar. Louvain: Peeters.

Moscati, Sabatino, Anton Spitaler, Edward Ullendorff, and Wolfram von Soden. 1964. An introduction to the comparative grammar of the Semitic languages, ed. by S. Moscati. Wiesbaden: Harrassowitz.

Nöldeke, Theodor. 1889. Semitic languages. Encyclopædia Britannica, 9th edition, revised 11th edition, 1911.

von Soden, Wolfram. 1952. Grundriss der akkadischen Grammatik [Foundations of Akkadian grammar]. Rome: Pontifical Biblical Institute, 3rd edition, 1995.

————. 1958–1981. Akkadisches Handwörterbuch [Akkadian desk dictionary], 3 vols. Wiesbaden: Harrassowitz.

Wright, William. 1890. Lectures on the comparative grammar of the Semitic languages. Cambridge: Cambridge University Press.

Zimmern, Heinrich. 1898. Vergleichende Grammatik der semitischen Sprachen [Comparative grammar of the Semitic languages]. Berlin: Reuther & Reichard.

PETER T. DANIELS

See also **Afroasiatic; Arabic; Aramaic; Philology**

Serbo-Croatian and South Slavic Languages

The South Slavic languages, Serbo-Croatian, Slovene, Bulgarian, and Macedonian, descend from Slavic dialects that were brought to the sub-Alpine and Balkan regions of southwestern Europe c. 500 CE by waves of westward migration along and across the Danube, Drava, and Sava river systems. In their new territory, the South Slavs encountered and undoubtedly mixed with Latin-speaking peoples, probably descendants of older Indo-European-speaking peoples, for example, Illyrian and Thracian.

The exact relationships among the dialects at the time of settlement are uncertain, but it is not the case that there were already nascent Slovene, Serbo-Croatian, Macedonian, and Bulgarian dialects. Rather, these formed over the subsequent millennium. The South Slavic group may now be defined by its geographical discontinuity to the remainder of the Slavic-speaking world. To the north, Slovene is bounded by Friulian and Italian in Italy, by German in Austria, and by Hungarian in Hungary. Croatian and Serbian are also bounded by Hungarian and Romanian (Romania). Bulgarian is bounded by Romanian,

which, together with the Black Sea, separates South Slavic from Ukrainian. Within the South Slavic branch, two subgroups are distinguished: Western South Slavic, constituted by Slovene and Serbo-Croatian, and Eastern South Slavic, constituted by Macedonian and Bulgarian. The languages are also divided along cultural and religious lines: Slovene and Croatian are spoken predominantly by Catholics, whereas Serbian, Macedonian, and Bulgarian are spoken by Eastern Orthodox Christians.

These divisions have determined the choice of alphabet, Latin being chosen in Catholic areas, and Cyrillic (a modified variety of the Greek alphabet) in Eastern Orthodox areas. Bosnia, which has been religiously and ethnically mixed and also includes a significant Muslim population, had vacillated among different alphabets. Since the disintegration of Yugoslavia, the standard Bosnian of Muslims is written in the Latin alphabet, whereas the Bosnian Serbs use Cyrillic. As with most Indo-European languages, the South Slavic group is characterized by many grammatical endings, with nouns and verbs changing form

depending on their position in the sentence or on their function as subjects or objects, singulars or plurals. Slovene and Serbo-Croatian go with the rest of the Slavic-speaking world in having preserved most of these endings in nouns, but verbs have become somewhat simplified. Macedonian and Bulgarian have the opposite: simplified nouns but more complicated verbs.

Serbo-Croatian is spoken by approximately 16 million people. It is the state language of the Republic of Croatia (where it is called Croatian), Bosnia and Herzegovina (where it is called Bosnian), and Serbia and Montenegro (where it is called Serbian); minority speakers are also found in Italy, Hungary, Austria, Romania, Bulgaria, and Macedonia.

The Serbo-Croatian standard was formed in the nineteenth century as a compromise between Serbs and Croats, whose major dialect divisions and corresponding divergent literary traditions, particularly in the Croatian case, had fostered disunity. The Čakavian and Kajkavian dialects, both spoken in Croatian ethnic territory, and which had developed into sophisticated literary vehicles during the Renaissance and Reformation, respectively, were abandoned as models for the standard language in favor of the Štokavian dialect, spoken in Croatia and all of Serbia, as well as in Bosnia and Herzegovina and Montenegro. In Serbia, the new Štokavian-based standard replaced the artificial Slaveno-Serbian literary language, which was based largely on Old Church Slavic. The compromise, which was engineered by intellectuals around the Croat Ljudevit Gaj and the Serb Vuk Karadžić, was codified in the Literary Agreement of 1850. The standard had two varieties, the Croatian (or Western), written in a modified Latin alphabet, and the Serbian (or Eastern), written in a modified Cyrillic. This standard persisted officially as the language of the Croats, Serbs, and (Bosnian and Sandžakian) Muslims, as well as the de facto lingua franca of Yugoslavia, until the disintegration of the state in 1991. Since then, separate Croatian, Serbian, and Bosnian state languages (the latter using the same alphabet as Croatian and having a relatively higher number of Turkish and other Islamic cultural borrowings) have been cultivated, each continuing from their inherited Štokavian-based precursor; all three standard languages remain almost completely mutually intelligible. (For this reason, 'Serbo-Croatian' persists as a linguistically valid term, referring to the speech territory and the common base of the separate language collectively. However, it is no longer considered an acceptable term to most lay speakers or the governments of the successor states.) Other regional movements, including notably a Montenegrin one, suggest the possibility of forming further standard languages in the future.

The Serbo-Croatian speech territory is characterized by three distinct dialect areas, each labeled by both professionals and the laity by the word meaning 'what'. A transitional zone called the Torlak group displays features of both Štokavian and neighboring Macedonian and is thus arguably within the scope of the Balkan *Sprachbund*, an area of linguistic convergence among distantly related or even unrelated languages caused by long-term contact, which also includes Albanian, Aromanian, Greek, Romanian, Romany, and, to some extent, Turkish.

Generally speaking, linguists' attention has been drawn to Serbo-Croatian (as well as Slovene), especially for its phonological (sound pattern) features, high degree of dialect variation, and preservation of key archaisms that aid in the reconstruction of Proto-Slavic, the prehistoric language thought to have been spoken by all Slavs before 500 CE. Standard Slovene and Serbo-Croatian, as reflected in many of their dialects, contrast long and short vowels, and, along with stress, have rising and falling tones (similar to Chinese), e.g. Slovene *brá:t(i)* 'to read' (long low pitch), *brà:t* 'to go read' (long high pitch), and *bràt* 'brother' (short high pitch). Other features are of interest, particularly word and sentence structure, e.g. Serbo-Croatian has begun to simplify its nouns, as has occurred more radically in Macedonian and Bulgarian, by reducing the number of grammatical endings ('cases'), especially in the plural.

Structurally, Slovene is closest to Serbo-Croatian and is spoken by approximately two million people, largely in the Republic of Slovenia, where it is the primary official language (alongside regionally official Italian and Hungarian). It is also spoken by significant minorities in neighboring Italy, Austria, and Hungary.

Modern standard Slovene, which began its development with the religious translations of the Protestant Primus Truber (Primož Trubar in Slovene) in the mid-sixteenth century, was established in largely its current form toward the end of the nineteenth century. It is based on the urban speech of the capital, Ljubljana, and the surrounding central dialects, although it also has features selected from its highly variegated dialects. It is written in a modified variety of the Latin alphabet, similar to Croatian.

With its relatively small speech territory, Slovene has seven dialect bases and greater internal differentiation than any of the South Slavic languages. Speakers from the most extreme dialects (e.g. Rezija, Prekmurje) generally cannot be understood by standard speakers. Slovene preserves archaic features that have been lost in Serbo-Croatian. For example, it distinguishes not just singular and plural but also dual

number (*pogovarjava se* 'we two are conversing'); it makes the future tense with an auxiliary verb and a participle (*bom sedela* 'I shall sit'); and it preserves a special 'supine' form of the verb that signals intention (*kupovat bom šel* 'I shall go to shop'). In contrast to Serbo-Croatian, Slovene has a relatively significant number of borrowings from German (e.g. *farba* 'color' from *Farbe*), Italian (*fant* 'boy' from *fante*), and Friulian (*križ* 'cross' from a seventh-century Friulian form *krože*).

Bulgarian is spoken by approximately nine million people, predominantly in the Republic of Bulgaria, where it is the primary state language, as well as by minority speakers in Serbia and Macedonia. Structurally, Bulgarian is closest in type to Macedonian.

Modern Bulgarian dates to the seventeenth century and developed substantially into its current form in the middle of the nineteenth century. It is based on the Tărnovo dialect of northeastern Bulgaria, but with elements from various dialect areas. Medieval varieties of Bulgarian served as the primary examples of Slavic writing, with prominent writing centers located in Preslav and Tărnovo. Modern Bulgarian is written in a modified variety of Cyrillic.

Macedonian is spoken by approximately two million people, primarily in the Republic of Macedonia. Significant groups of Macedonian speakers are also found in northern Greece, western Bulgaria, Serbia, and in some villages in Albania.

Macedonian was codified as a standard language in 1944, although the beginnings of the contemporary language may be traced to the middle of the nineteenth century. Macedonian is written in a modified variety of the Cyrillic alphabet. The language of the Macedonian speech territory can be traced back organically to the speech that gave rise to the first Slavic written language in the ninth century CE, known today as Old Church Slavic.

Linguists have tended to concentrate on the structure of Macedonian and Bulgarian words and their relationship to syntax and meaning, as well as the interaction of the languages with others in the Balkan linguistic convergence area (or *Sprachbund*). For the period from the tenth to the twelfth centuries, the textual evidence of Proto-Macedonian and Bulgarian is important for the earliest body of attestations of Slavic in general, known as the canonical period of Old Church Slavic. For this reason, Indo-Europeanists have made substantial use of older Macedo-Bulgarian material.

Because of their participation in the convergence area, Macedonian and Bulgarian display features not found elsewhere in the Slavic-speaking world. For example, the category of definiteness is marked by the presence (vs. absence) of an article after the first member of a noun phrase, e.g. Macedonian *Ja vidov zhenata* 'I saw the [a certain] woman' vs. *Vidov zhena* 'I saw a woman'; Serbo-Croatian makes no such distinction, having only *Vidjela sam ženu* 'I saw the/a woman.' A distinction expressed by choices among alternative verb forms is made between witnessed and nonwitnessed events, e.g. Bulgarian *Toj napisa pismoto* 'he wrote the letter [I know so because I saw him do it]' vs. *Toj napisal pismoto* 'he wrote the letter [so it is said—I did not see him do it]'; Serbo-Croatian makes no such distinction, having only *Napisao je pismo* 'He wrote the/a letter.' The inherited infinitive has been lost and replaced by a subordinate clause, e.g. Bulgarian *Iskam da otida na mač* 'I want to go [literally 'that I go'] to a game' vs. Serbo-Croatian *Hoću ići na utakmicu* 'I want to go to a game.' The origin of such convergence features is much debated: they may be a continuation of structures from languages that have disappeared (substratum languages)—Illyrian and Thracian—or a result of language contact itself and diffusion of linguistic features, although the working of both explanations together is not excluded.

The South Slavic languages represent a picture of great diversity among the Slavic languages, and, because they are located at a crossroads of European languages and cultures, they have been affected by contacts with numerous languages. The volatile political fortunes of the region promise to push the development of the languages, especially the newly differentiated Bosnian, Croatian, and Serbian, toward ever greater diversity.

References

Browne, Wayles. 1993. Serbo-Croat. In Comrie and Corbett.

Comrie, Bernard, and Greville G. Corbett (eds.) 1993. The Slavonic languages. London: Routledge.

Friedman, Victor A. 1993. Macedonian. In Comrie and Corbett.

Greenberg, Marc L. 2000. A historical phonology of the Slovene language (= Historical phonology of the Slavic languages XIII). Heidelberg: Carl Winter Universitätsverlag.

Huntley, David. 1993. Old Church Slavonic. In Comrie and Corbett.

Ivić, Pavle. 1958. Die serbokroatischen Dialekte: Ihre Struktur und Entwicklung. The Hague: Mouton & Co.

Katičić, Radoslav. 1976. Ancient languages of the Balkans. The Hague: Mouton.

Koneski, Blaže. 1983. A historical phonology of the Macedonian language (= Historical phonology of the Slavic languages XII), transl. by V. A. Friedman. Heidelberg: Carl Winter Universitätsverlag.

Lehiste, Ilse, and Pavle Ivić. 1986. Word and sentence prosody in Serbocroatian. Cambridge, MA: The MIT Press.

Lencek, Rado L. 1982. The structure and history of the Slovene language. Columbus: Slavica.

Priestly, T. M. S. 1993. Slovene. In Comrie and Corbett.

Scatton, Ernest A. 1993. Bulgarian. In Comrie and Corbett.

MARC L. GREENBERG

Serial Verb Constructions

Some languages—especially those of West Africa, Southeast Asia, and New Guinea—allow several verbs to appear within a clause to express a single meaning, as illustrated from Kalam (Papua New Guinea), where 'massage' is expressed by the following string of verbs:

pk	*wyk*	*d*	*ap*	*tan*
'strike	rub	hold	come	ascend

d	*ap*	*yap*	*g-*
hold	come	descend	do'

Serial verbs are so prevalent in Kalam that most verbal meanings can be expressed by combining fewer than 100 verb roots.

In other languages, just two or three verbs are serialized. In Paamese (Vanuatu), *vul-* means 'break', but this verb is seldom used, the strong preference being to serialize it with another verb indicating the precise activity that caused the breakage, for example:

Ni-matil	*vul*	*a:i.*
1SING.FUTURE-sleep	break	plank

'I will sleep on the plank, thereby breaking it.'

These constructions represent single clauses, in contrast to both subordinate and coordinate multi-clause constructions. However, it is difficult to clearly and consistently define the term 'serial verb', and there is also the related problem of the wide variation in terminology used to refer to structurally similar or historically related phenomena in different languages. Different writers have referred to 'verbids', 'locative verbs', 'post verbs', 'postverbal prepositions', 'locative copulas', and 'coverbs' to refer to essentially the same patterns in West African languages. Many are not explicit about what they mean by serial verbs, some treating any verb–verb sequences as serial verbs as long as the second verb is not marked as an infinitive, e.g. *Go get the book*. Sebba (1987) provides a more explicit set of criteria for recognizing serial verbs:

(1) Both verbs must be able to function independently within a clause as verbs in their own right.
(2) Both must be interpreted as having the same categories of tense–aspect–mood.
(3) There must be no marking of a clause boundary.
(4) There should be no conjunction appearing between the two.

Some insist that serial verbs must be distinguished from verbal compounding. However, the literature on nominal compounding is replete with statements of difficulties in distinguishing between phrases such as *wood turner* and words such as *woodlice*. Verbal compounding is no less problematic, and any attempt to categorize verbal compounds in many languages typically runs up against the problem of deciding where the boundary between compounding and serialization lies.

However, this has the potential to divert attention away from the real issues. The relevant structures could simply be regarded as complex verbs composed of more than one grammatical element, avoiding a distinction between verbal compounding and verb serialization. The two, in fact, exhibit a variety of similarities crosslinguistically, with the only real difference being that we are dealing with either words or phrases.

Verb serialization can be distinguished from subordination and coordination in that there should be no sentence-level connectives between the various verbal elements. If there is any evidence of subordinators (such as *that* or *which*) or coordinators (such as *and* or *or*)—even if only as an optional variant, as we find with English *Go (and/to) get the book*—then we are dealing with a complex sentence rather than a single clause.

There is considerable diversity within the category of serial verbs among different languages and even within individual languages, with some kinds of serial verb constructions behaving in ways that more closely resemble verbal compounds, and others sharing features with subordinate or coordinate constructions.

Many verb-final languages have clause chains that are structurally midway between serial verbs and complex sentences. There may be extensive juxtaposition of verb phrases without any subordinating or coordinating markers, but with considerable contextual ellipsis of shared arguments. This means that we are dealing with a structural continuum as follows: verbal compounds > serial verbs > clause chains > subordinate clauses > coordinate clauses. It is probably wishful thinking to assume that we can produce a universally applicable definition of verb serialization: rather than a separate universal category, serialization is more accurately characterized as a syndrome of features and phenomena.

References

Alsina, Alex, Joan Bresnan, and Peter Sells (eds.) 1997. Complex predicates. Stanford: Centre for the Study of Language and Information.

Baker, Mark C. 1989. Object sharing and projection in serial verb constructions. Linguistic Inquiry 20.

Crowley, Terry. 2002. Serial verbs in Oceanic: a descriptive typology. Oxford: Oxford University Press.

Durie, Mark. 1997. Grammatical structures in verb serialisation. In Alsina, Bresnan, and Sells.

Lord, Carole. 1993. Historical change in serial verb constructions. Amsterdam: John Benjamins.

Pawley, Andrew. 1993. A language which defies description by ordinary means. The role of theory in language description. ed. by William A. Foley. Berlin: Mouton de Gruyter.

Sebba, Mark. 1987. The syntax of serial verbs: an investigation into serialisation in Sranan and other languages. Amsterdam: John Benjamins Publishing Company.

TERRY CROWLEY

Sign Relationships

A signifier, or 'sign' for short, is any *x* that stands for or represents some *y*, that calls some *y* to mind (or is thought of as capable of so doing). For example, a south-pointing weathervane can be a 'sign' that the local wind is a southerly, if so interpreted by a well-disposed observer (Watt 1993:432). By the same token, the word 'unicorn' is a sign of, or signifies, a mythical equine creature with a single helical horn peculiarly susceptible to the importunities of virgins. In the latter case, 'signifies' can be replaced by 'means', but this is not always so. For instance, it would be odd to say that the letter 'A', which signifies the set of sounds comprising the phoneme /a/, 'means' those sounds. And whereas it is safe to say that the word 'thwirk', which has never existed till now, has no meaning at all in the ordinary way, it still signifies something (e.g. that its creator knows how to observe the rules of English phonology). Hence, 'signifies' is the more general term, since 'means' will sometimes seem misused when neither *y* ('a mythical equine. . .') nor *x* ('unicorn') is a linguistic expression.

All living things, from microbes to elephants, and including plants and humans, process signs of one sort or another, whether they be words and sentences or, for plants, the signals of chemicals and light (Sebeok 1989:12–24). Restricting our purview to humans, however, it is safe to say that every signification, every tying of some signified *y* to some signifier *x*, must at a minimum happen in the mind of a human recipient who perceives the *x* and ties a *y* to it. Some other human may have sent the *x* in the hope of its being properly interpreted; but many *x*'s are interpreted in such a way as to defeat any such hope. And many *x*'s are not sent at all, by anybody anywhere. Should a sender use the word 'livid', he or she is likely to find it misinterpreted as meaning 'furious'; when an earthquake is interpreted as signifying God's wrath, this is not owing to its having aimed at that result (or at any other). In sum, all that is necessary for an act of signification to take place is one mind's perceiving some *x*, whatever its source, and tying it to some *y*, however legitimately.

Language is the supreme signifier. Its morphemes, words, phrases, clauses, sentences, paragraphs, and whole discourses are almost entirely emitted in the hope of signifying, whether to ourselves or to others; and we listen and read in the wan hope of being informed. *Homo sapiens* is, above all, *homo signifex* (*homo* the sign-maker).

Signs of whatever sort are often bound into a set, a system, characterized by its members' widespread sharing of attributes. Many such attributes are unconsciously known to the signs' users, since often they can reliably judge whether or not a proposed new member of the set is sufficiently like the old ones (i.e. shares enough of their attributes) to win entry. Some such judgments are quite subtle. For instance, a speaker of English will accept 'thwill' as a possible (or unknown) English word, while rejecting 'thpill' (despite the currency of 'thpill' in the speech of lispers); users of the Romani English alphabet are more likely to accept '0' as a possible new letter than '0' (Jameson 1994). Besides being related to others within their set, signs are also related to what they signify. Thus, a sign may be physically associated with what it signifies, as the weathervane is to the wind that actuates it; or it may be perceived as being physically similar to what it signifies, as the word 'whinny' is to the sound made by a horse; or it may have neither of those ties, being purely arbitrary or conventional, as the word 'horse' is. Thus, in Charles S. Peirce's well-known terms, a sign can be, respectively, 'indexical', 'iconic', or 'symbolic' (Peirce 1935–1966:2.275). Many signs in their

actual occurrence, including words or morphemes, partake of two or even three of these qualities: the stick figure representing 'MAN', for instance, conveys 'Here's a men's room' only when it is on or near an expectably placed door, thus being indexical; it looks a little like a man, thus being also iconic; but it is still highly abstracted or conventional, thus being symbolic as well. And, turning now to words, since 'whinny' is but a poor imitation, it is also partly arbitrary, i.e. symbolic. Even the structure and superficial word order of sentences, although in large part arbitrary (which is why different languages can order things differently), can even then be said to be also somewhat iconic, as when, in English, the order of cardinal sentential elements seems to mirror a feeling that one is better understood when new or indefinite information is withheld till after a given or definite context has been established (Halliday 1967; Clark and Haviland 1977). (Thus, although ordinarily an active sentence is more felicitous than a passive one, this preference is reversed when the underlying subject is indefinite and the object is definite: the passive 'The General's been hit by an arrow' is likely to strike most speakers as more natural than the active 'An arrow's hit the General'.)

The signs habitually used by humans are also, naturally enough, related to their human users. The sounds of a natural language are those the human vocal tract can easily make and the human ear can distinguish, and the letters of the Romani English alphabet are those the human hand + pen (the 'scribal tract') can easily draw and the human eye can discriminate—an obvious point, but one not noted enough.

The signs of language differ but little, at the lower levels (phonemes; morphemes), from the signs of some other systems, either in complexity or even in sheer numbers. Morphemes and the cattle brands of the American West, for instance, have comparable morphologies. On the other hand, the signs of language at the higher levels (phrases, clauses, sentences, discourses) differ in kind from those of any other sign system known, both in their formal complexity and, of course, in their number, which is unbounded. However, despite their differences, all signs, of whatsoever kind, lend themselves to description in terms of how they are related to what they signify, to each other, and to those who send and receive their physical manifestations and interpret them.

References

Clark, H.H., and S.B. Haviland. 1977. Comprehension and the given-new contract. Discourse production and comprehension, ed. by R. Freedle. Norwood, NJ: Ablex.

Halliday, M.A.K. 1967. Notes on transitivity and theme in English: ll. Journal of Linguistics 3. 199–244.

Jameson, K. 1994. Empirical methods for evaluating generative semiotic models: an application to the Roman majuscules. Writing systems and cognition: perspectives from psychology, physiology, linguistics, and semiotics, ed. by W.C. Watt. Dordrecht: Kluwer.

Peirce, C.S. 1935–1966. Collected papers of Charles Sanders Peirce, ed. by C. Hartshome, P. Weiss and A.W. Burks (eds.). Cambridge, MA: Harvard University Press.

Sebeok, T.A. 1989. The sign & its masters. Lanham, MD: University Press of America.

Watt, W.C. 1993. Signification and its discontents. Semiotica 97. 427–37.

WILLIAM C. WATT

Signed Languages

Signed languages are a class of languages in which speech is replaced by manual and facial gesture and auditory reception is replaced by vision. Natural signed languages have evolved within the boundaries of subcultures of deaf communities throughout the world. The term 'signed language' is often used more generally to include alternate sign languages, manual sign systems developed in certain hearing communities as alternatives to spoken language that serve social, religious, or occupational functions. The term is also used to describe manual language codes, sign systems invented by educators to represent spoken languages in manual gestures. The use of the term here will be restricted to those natural signed languages that evolved within deaf communities. A natural sign language is learned as a first language by children in the community, is not related to the spoken language of the broader culture, and possesses the complexity and types of structural features found in spoken languages.

The use of manual signs dates back at least to the early Greeks; Plato mentioned the use of signs by deaf individuals in the *Cratylus*. Signed languages are, however, a recent addition to the study of language and language structure. Serious study of signed

languages began only in the last half of the twentieth century. Prior to that time, signs used in deaf communities were believed to be rooted in a biological gestural system that was universal and, to a great extent, iconic; and signed communication was thought to lack the complexity of spoken languages. Two events occurred in the mid-1900s that changed the status of signed languages. The first systematic analysis of a signed language became available in 1960. In that year, William Stokoe published a monograph on American Sign Language (ASL). Stokoe demonstrated that the signs of ASL possessed an abstract sublexical structure analogous to the phonological structure of words in spoken languages. Five years later, he and two colleagues (Dorothy Casterline and Carl Croneberg) published the first dictionary of a signed language (ASL) based on linguistic principles. In this same time period, rapidly developing video and film technology provided an economical way to capture and preserve dynamic properties of signed languages for analysis. These two developments, video techniques to record basic language data and Stokoe's demonstration that signed languages were subject to the same analytical techniques as spoken languages, provided the basis for the linguistic study of signed languages throughout the world.

Distribution

Currently, at least preliminary descriptions exist for well over 100 signed languages, spanning the major geographical regions of Africa, Asia, Australasia, Europe, and North and South America. The names of these languages typically reflect the geographical area(s) where they are used (e.g. Kenyan Sign Language, Japanese Sign Language, Australian Sign Language or Auslan, Italian Sign Language, ASL, and Brazilian Sign Language, respectively). Many signed languages developed independently within the deaf communities of somewhat isolated regions. Providence Island Sign Language, for example, is the primary language of fewer than two dozen deaf people and a second language for some hearing people on Providence Island, Columbia, a small island off the coast of Nicaragua. Providence Island Sign Language developed independently of the spoken language (English and English Creole) of the island, although deaf residents had no exposure to other signed languages. Signed languages have also developed through processes of migration and cultural contact. With time, historical change produces distinctly different signed languages. For example, three distinct signed languages have been identified in Vietnam: Ho Chi Minh City Sign Language, Hanoi Sign Language, and Haiphong Sign Language. Available evidence of

partial intelligibility indicates that these three languages belong to the same language family.

In many cases, historical and social forces have influenced deaf communities differently from hearing communities in the same regions. Geographical boundaries sometimes follow those for spoken languages, but they often do not. ASL and British Sign Language are not mutually intelligible, for example, although English is the spoken language of both geographic regions. Instead, there are some similarities between ASL and French Sign Language (FSL) because modern ASL is derived from contact between one or more indigenous sign languages in the United States and FSL, brought to the United States in the early 1800s.

Circumstances promoting the emergence of a new signed language have recently provided linguists with the opportunity to examine first hand the forces driving the creation of a new language. The birth of Nicaraguan Sign Language has been extensively documented over the past 25 years. When the Sandinistas came into power in Nicaragua in 1979, formerly isolated deaf children from all parts of the country were brought together in a new residential school for deaf children. The children who arrived possessed only rudimentary idiosyncratic home sign systems. Home sign systems are created by deaf children who live in isolation and have little or no conventional language input. Upon arriving at the school, the Nicaraguan school children immediately began to use signs from their home sign systems with each other. The result was the development of a pidgin language between home sign systems. Exposure of the next generation of school children to the pidgin resulted in the development of a more complex signed language through a gradual process of creolization. The resulting language is the language now called Nicaraguan Sign Language.

Language Structure: Universals and Modality Effects

Abstractness is a defining quality of human language. The addition of signed languages to the inventory of world languages has raised the level of abstractness. Languages are no longer restricted to the vocal–auditory channel. Elimination of this one-time universal appears to have no effect on other proposed universals. Signed languages satisfy most proposed universals, and, where some do not, spoken language counterexamples have also been discovered. Signed languages allow complexity of expression equivalent to that of spoken languages with complexity of grammar that is equivalent to grammars of spoken languages. ASL, for example, shares features with spoken Navajo and Japanese. The form and complexity of grammatical properties in signed languages do not differ from those

of spoken languages, but the means by which those grammatical properties are implemented differ by modality.

Stokoe's analysis of the sublexical structure of ASL signs provided the first example of how a universal grammatical property can be implemented in the manual modality. Stokoe demonstrated that ASL signs could be described by three phonological categories of arbitrary components: 19 handshapes, 12 locations, and 24 movement patterns. Minimal pairs of signs are identical except for a change of one component from one phonological category. Orientation of the hand is also needed in some cases to discriminate certain minimal pairs of ASL signs. Components from the phonological categories were originally thought to occur simultaneously in sign formation. Later analyses, however, have demonstrated that signs, like spoken words, have linear segmental structure. Analyses indicate that sign languages throughout the world use similar components at the phonological level, but the specific inventory of handshapes, locations, and movements differs from language to language.

The use of the gestural–visual channel provides both advantages and constraints for the development of grammatical mechanisms in signed languages. The use of manual articulators results in gross motor movements that take more time to produce than the fine motor movements involved in speech. Time constraints on linear sequencing are complemented, however, by the availability of two articulators, the greater simultaneity of components in word formation in signed languages, the use of three-dimensional space, and the coordinated use of nonmanual and manual components. Some language processes based in the gestural–visual channel appear to be universal to signed languages, whereas others are language specific or are found in a subset of signed languages.

All signed languages analyzed to date possess complex systems for the creation of multimorphemic signs. In spoken languages, complex words are typically created by processes of concatenation. Complex word formation processes differ for signed languages in that simultaneous morphological components take the place of linear strings of affixes. Temporal aspect inflection is carried out through changes in the movement characteristics of sign stems in many signed languages, including ASL, British Sign Language, and Swedish Sign Language. Derivational processes creating nouns from verbs in ASL also involve changes in the movement patterns of the verb forms. In general, the linear affixation processes prominent in spoken languages are rare or nonexistent in signed languages. Instead, these examples of inflectional and derivational morphology involve the simultaneous occurrence of a base sign and morphological markings.

Classifier constructions are ubiquitous in signed languages. Classifier systems take advantage of the spatial medium to express types of motion through space and position in space, as well as stative-descriptive and handling information. Complex predicate classifier constructions are universal to all signed languages analyzed to date. They also exist in evolving languages like Nicaraguan Sign Language and in idiosyncratic home sign systems. In classifier predicates, handshapes that function as classifier morphemes combine with different types of movement morphemes to form complex predicates. Categories of handshape and movement morphemes found in both ASL and Danish Sign Language (taken from Emmorey 2002) illustrate the properties of signed language classifier systems.

Categories of handshape and movement morphemes for ASL and Danish Sign Language include the following:

Handshape Morphemes

Whole entity classifiers: refer to an object as a whole
Handling/instrument classifiers: represent a whole entity, but imply an agent handling the entity
Limb classifiers: represent limbs of animals and humans
Extension and surface classifiers: represent depth and width of an object or width of a surface

Movement Morphemes

Position morphemes: indicate 'to be located', posture, or a change in object orientation
Motion morphemes: indicate movement along a path
Manner morphemes: indicate manner (e.g. speed) of motion without specifying a path
Extension morphemes: depict the outline or configuration of an entity or mass

Each language imposes certain constraints on possible morpheme combinations. Whole entity classifiers combine with extension movement morphemes to indicate spatial arrangement of multiple objects. For example, the ASL handshape classifier for vehicles combined with the linear extension movement morpheme indicates 'cars in a row'. The instrument handshape classifier used to represent 'knife' could not, however, be substituted for the whole entity 'vehicle' classifier to indicate 'a row of knives'.

Signed languages, but not spoken languages, allow for certain types of simultaneous constructions. These constructions involve the use of the two hands to express two different predicates at the same time. In ASL, for example, a simultaneous construction can be used to express, 'The cat jumps on a shelf'. In this construction, the nondominant hand predicate indicates the background element, 'the shelf'. The handshape classifier for narrow surfaces (a flat hand) is combined with the 'hold' morpheme, a static movement morpheme

that indicates static motion or holds the motion of an entity in place for purposes of the construction. The dominant hand predicate combines a classifier handshape used for small animals with an 'arc' motion morpheme. The end of the arc motion makes contact with the nondominate hand, which is located above the level of the starting point of the arc motion.

Manual signs are normally expressed in a signing space that extends vertically from just below the waist to the top of the head and horizontally from just over the signer's shoulder outward a little more than a foot. Lines defined in this space are used to express different forms of temporal reference. The abstract concept of time is overlaid on these time lines. The use of time lines to convey temporal information is found in languages as diverse as ASL, Argentine Sign Language, Australian Sign Language, British Sign Language, Jordanian Sign Language, and the Sign Language of the Netherlands. At least two of these time lines may be universal to signed langauges. One of these time lines is related to immediate context. This time line extends from the top of the dominant shoulder outward, perpendicular to the body. The signer's body provides the reference point for 'the present', with the past behind the signer and the future in front. Temporal adverbs make use of this time line. For example, the sign for YESTERDAY moves backward from the side of the cheek, and TOMORROW moves forward. Adverbial modification of other signs is used to express the past or future. In ASL, for example, the signs for days of the week are made near the body. These signs can be modified to indicate the next occurrence of that day by changing the location of the sign to a point forward on the time line (e.g. MONDAY modified to represent NEXT-MONDAY). A second time line represents succession or duration of time. This time line extends horizontally, parallel to the front of the signer's body. Sign movement from left to right along this time line establishes temporal reference in discourse from early to late in many signed languages. The British Sign Language sign CONTINUE uses this kind of movement to express the persistence of an event or state over a long period of time. Jordanian Sign Language also makes use of the sequential time line, but time from early to late is expressed in right to left movement.

Another common, probably universal, mechanism found in signed languages is the use of spatial location to refer back to something that was previously mentioned and to indicate which nouns in the sentence are subject and object, respectively (agreement). In signed languages, personal pronouns take the form of pointing gestures. Pointing has been co-opted from the gestural realm to form the basis for linguistic pronominal systems in signed languages. Points in signed languages are categorized into handshape, location, and movement components that are individually used to convey different grammatical distinctions that may include person, number, and case. Pointing is used to refer to entities, whether present or not. When the referent is present, the signer points to the actual position of the referent. When the referent is absent, the signer establishes a spatial location for the referent and points toward that spatial location. Signed languages have conventionalized ways of assigning spatial locations for different types of referents.

All signed languages examined to date use modifications of the verb to mark grammatical agreement with subject and object for some classes of verbs. Agreement is marked through modification of the sign's hand orientation and location and direction of the movement. The movement of these verbs is directed from the spatial location marking the subject toward the spatial location marking the object. In Italian Sign Language, for example, the sign for TEACH is composed of an initial handshape that is closed such that the fingertips touch the tip of the thumb. The hand opens to a flat handshape with the fingers spreading as the movement of the sign is produced. In the sentence 'I teach you,' the movement of the sign begins near the signer's dominant shoulder and moves outward toward the spatial location of the addressee. The front/palm of the hand faces the addressee throughout the production of the sign. The location, movement, and hand orientation of the sign TEACH are modified to mark subject and object agreement in the sentence, 'Elena teaches Louis.' If the spatial position marking Elena is to the signer's right and the spatial position marking Louis is to the left, the movement of the verb TEACH is from right to left, beginning at the spatial location marking the subject, Elena, and moving toward the spatial position marking the object, Louis. The hand is oriented toward the spatial location marking Louis as the handshape changes from closed to open.

In addition to the manual channel, signed languages make use of a second, nonmanual, channel to convey grammatical information. Facial expression and changes in head and body position are used to provide distinct linguistic contrasts simultaneously with the production of manual signs. Nonmanual markers are used grammatically to distinguish similar structures across signed languages. These structures include negation, Wh- questions, Yes–No questions, conditionals, relative clauses, and rhetorical questions. In general, the duration of the nonmanual marker indicates to which part of the sentence it applies. The nonmanual component used to mark negation, for example, typically includes a specific facial expression with accompanying head movement. The grammatical facial expression is maintained for the entire duration that it takes to produce the phrase being

negated. Nonmanual components appear to be similar in both form and function across signed languages.

Neural Systems for Language

In the same way that the form of language is not dependent on language modality, the neural systems underlying language processing are not dependent on language modality. Brain-damaged signers display similar types of error patterns as brain-damaged speakers. Language is typically processed in the left hemisphere for both signed and spoken language and specific types of language disorders are localized in the same neuroanatomical structures of the left hemisphere. Classic patterns of Broca's and Wernicke's aphasia have been described for deaf signers. Prosodic disturbances and reduced, 'telegraphic' syntactic constructions are associated with damage to Broca's area of the left hemisphere for both spoken and signed language. Errors in word or phoneme choice and the production of jargon (grammatically correct but unintelligible sentences) are associated with left hemisphere damage in Wernicke's area for both spoken and signed language. Wernicke's area is also involved in language reception. Specific sites in Wernicke's area have been associated with the processing of phonetic–syllabic units, independent of language modality. The left hemisphere appears to be specialized for language, independent of modality. The abstract, modality-independent nature of language, both in language form and in functional neuroanatomy, has rekindled debate about the origin of language.

References

Ahlgren, Inger, Brita Bergman, and Mary Brennan (eds.) 1994. Perspectives on sign language usage, Vols. 1–2. Durham: International Sign Linguistics Association, University of Durham.

Brentari, Diane. 1998. A prosodic model of sign language processing. Cambridge, MA: MIT Press.

Emmorey, Karen. 2002. Language, cognition, and the brain. Mahwah, NJ and London: Lawrence Erlbaum Associates.

———. (ed.) 2003. Perspectives on classifier constructions in sign languages. Mahwah, NJ and London: Lawrence Erlbaum Associates.

Klima, Edward S., and Ursula Bellugi. 1979. The signs of language. Cambridge, MA: Harvard University Press.

Meier, Richard P., Kearsy Cormier, and David Quinto-Pozos (eds.) 2002. Modality and structure in signed and spoken language. Cambridge: Cambridge University Press.

Neidle, Carol, Judy Kegl, Dawn MacLaughlin, Benjamin Bahan, and Robert G. Lee. 2000. The syntax of American sign language: functional categories and hierarchical structure. Cambridge, MA: MIT Press.

Poizner, Howard, Edward S. Klima, and Ursula Bellugi. 1987. What the hands reveal about the brain. Cambridge, MA: MIT Press.

Stokoe Jr., William C., Dorothy C. Casterline, and Carl G. Croneberg. 1965. A dictionary of American Sign Language on linguistic principles. Washington, DC: Gallaudet College Press.

PATRICIA SIPLE

Sindhi

Sindhi is a language that originated in the Lower Indus Valley region of the Indian subcontinent, spoken by over 40 million people in present-day Pakistan and India and by a large diaspora community around the world. Sindhi belongs to the Indo-Aryan language family within Indo-European, and is classed with Kashmiri in the Northwestern subgroup. Sindhi is the primary language of the province of Sindh in Pakistan, and is spoken along the Indus River Valley stretching into the Thar Desert to the east, and bounded by the Sukkur Dam to the north, the Kirthar Mountain Range to the west, and the Arabian Sea to the southwest. To the south, the Sindhi region extends into the Rann of Kacch in India. While Sindhi is used exclusively or as the primary language in most rural areas within this region, it exists alongside Urdu and English in the urban centers of Sindh Province, including Karachi and Hyderabad, and alongside Hindi, Gujarati, and other regional languages in India. Sindhi is closely related to Siraiki, spoken in the north of Sindh Province, and to Kachhi, spoken to the south in the Kacch region of Gujarat in India.

Background

Sindhi shares many features in common with related Indo-Aryan languages. The sound inventory includes the distinctive voiced and voiceless aspirated obstruents at five places of articulation (labial, dental-alveolar, postalveolar, palato-alveolar, and velar), and a full set of paired long and short vowels, all of which can occur with nasalization. Common features in the morphology include number, gender, and case marking for nouns, a rich system of verb inflection, and a productive

process of compound verb formation. The basic syntactic structures of Sindhi are also common to Indo-Aryan, including a canonical Subject–Object–Verb (SOV) word order that is subject to permutation (i.e. so-called free word order), question formation with no preposing of question words (i.e. question words in situ), and dependent clauses involving the parallel relative–corelative construction.

Sindhi also has several features that differentiate it from other Indo-Aryan languages. Its consonant inventory includes four implosive stops, unique to Sindhi (and its close relatives such as Kacchi and Siraiki) among all Indo-European languages. Sindhi has retained from Sanskrit the full set of five phonemic nasal stops, but reduces all historical geminate consonants -CC- in the word-medial position to a singleton consonant -C-. Although the vowel inventory in Sindhi is common to Indo-Aryan, the status of the word-final short vowels is remarkable. Final short /i,u,a/ express grammatical information such as number, gender, and case on nouns, and yet they are produced with extremely short phonetic duration. These final vowels are typically not recognized by nonnative speakers, despite their important morpho-syntactic function, and are not retained in Sindhi words (e.g. proper names) that are adopted into Urdu, Hindi, or English. Although syllable structure in Sindhi is similar to other Indo-Aryan languages, Sindhi does not allow word-final consonants. Indeed, many Sindhi words that end in a final super-short vowel have cognate forms in related languages that end in a consonant. With respect to morphology, Sindhi differs from most Indo-Aryan languages in its use of pronominal clitics that attach to nouns, postpositions, and verbs. These clitics take the place of full noun phrases to express e.g. the possessor of a noun (*hatʰa-mi* 'my hands'), the complement of a postposition (*vəʈi-mi* 'belonging to me'), and with different clitic sets, the subject, object, or other complement of a verb (*ɓudʰaːrjãː-va* 'I shall tell you').

Linguistic studies of Sindhi are few in number, but include instrumental phonetic studies (Nihalani 1986, 1995), sociolinguistic and dialect studies (Rohra 1971, Bhugio 2001), and contemporary grammatical analysis (Khubchandani 1961). Published Sindhi grammars include Trumpp (1872), Grierson (1919), and Yegorova (1971).

Key Structural Features

Sound Inventory
The Sindhi sound inventory is shown below, and is notable for its inclusion of four implosive stops, which derive from Middle Indo-Aryan geminate voiced stops in medial position, and from singleton voiced stops in the initial position. Phonetic studies show that these

sounds are genuine glottalic ingressives (Nihalani 1986). The postalveolar consonants are apical and the palato-alveolar consonants are laminal; these are termed retroflex and palatal in traditional nomenclature. Sounds that are shown in parentheses in the chart below are either restricted to borrowings (/z, ʃ, x, ɣ/) or occur only as allophonic variants. For instance, the postalveolar rhotic tap [ɽ] is in complementary distribution with the homorganic stop [ɖ]. The aspirated sonorants /mʰ, nʰ, ŋʰ, lʰ ɽʰ,ʋʰ/ occur only in intervocalic position and are never distinguished from a sequence of sonorant + /h/, although speakers will describe them as single sounds, even though they are written as sequences in the Arabic orthography. The vowel inventory consists of the standard symmetrical Indo-Aryan system, comprising five pairs of long and short vowels. Phonetic values shown below are as identified by Nihalani (1995). In addition, all the vowels can occur with nasalization; the long nasal vowels have phonemic status, but nasalization on short vowels can be ascribed to the context of a following tautosyllabic nasal stop (see Table 1).

Morphophonology
Sindhi has a complex system of morphophonological vowel alternations that affect stem-final vowels in certain morphological constructions. These changes are highly idiosyncratic to the specific construction. For example, nouns are marked for plural number by modification of the stem-final thematic vowel alone or in combination with a plural suffix, resulting in plural endings marked variously by /-a, -aː, -ũː, -iũː/, or unmarked, depending on the gender class of the noun and on its lexically determined stem-final thematic vowel: *gʰar-a* 'houses', *pʰiːt-aː* 'wheels', *kʰaʈ-ũː* 'cots', *ɓili- ũ ː* 'cats', *raːti* 'nights'.

There are also very many verb stems that undergo irregular allomorphy in the formation of the Unspecified Perfective (i.e. simple past) tense. A comprehensive discussion of these alternations is provided in Trumpp (1872) and Grierson (1919).

Morphology

Nouns, Adjectives, and Pronouns
Nouns are classed by grammatical gender, and this classification determines the declension pattern for the marking of number and case, expressed primarily through stem vowel alternations. The noun stem ends in a thematic vowel, which for most nouns serves to mark the gender class, although there are numerous exceptional stems whose gender is not predictable on the basis of the final vowel. All nouns are marked for number and case through noun stem modification,

TABLE 1 Sindhi Consonant Inventory

	Labial	Dental	Alveolar	Post alveolar	Palatoalveolar	Velar	Glottal
Stop	p b	t d		ʈ ɖ		k g	
	pʰ bʰ	tʰ dʰ		ʈʰ ɖʰ		kʰ gʰ	
Implosive	ɓ		ɗ		ʄ	ɠ	
Affricate					tʃ tʃʰ		
					dʒ dʒʰ		
Nasal	m		n	ɳ	ɲ	ŋ	
	(mʰ)		(nʰ)	(ɳʰ)			
Fricative	f		s (z)		(ʃ)	(x ɣ)	h
Rhotic			r	(ɽ)			
				ɽʰ			
Lateral		l					
		(lʰ)					
Glide	ʋ				y		
	(ʋʰ)						

Sindhi Vowel Inventory

iː	[i]	eː	[e]	aː	[ɑ]	oː	[o]	uː	[u]
i	[ɪ]	e	[ɛ]	a	[ə]	o	[ɔ]	u	[ʊ]

involving a change in the thematic vowel, sometimes accompanied by suffixation. Patterns of noun stem modification that mark the plural and the nominative, oblique, ablative and vocative cases are illustrated in the example noun paradigms as shown in Table 2 for distinct thematic vowels, grouped by gender class.

The dative, ablative, comitative, and locative cases are marked through the use of a postposition following the noun in the oblique singular form. The genitive postposition has eight variants (of which only two are shown below); it is declined like an adjective and agrees in number and gender with the possessed noun, but bears the case specification of the possessor. There is no accusative postposition, and instead the nominative or dative form is used (depending on the animacy of subject and object). Also, the oblique case forms are used to mark ergative subjects (see Syntax section below).

Case/number marking with postpositions in example phrases with *gʰar-* 'house'

Dative	gʰara kʰeː	*'the house (direct object)'*
Ablative	gʰara kʰã̄ː	*'from (the direction of) the house'*
	gʰara mã̄ː	*'from in the house'*
	gʰara tã̄ː	*'from on the house'*
Comitative	gʰara sã̄ː	*'with the house'*
Locative	gʰara mẽ̄ː	*'in the house'*
	gʰara teː	*'at/on the house'*
	gʰara maɲdʒʰi	*'within the house'*

Genitive		
(masc.)	gʰara joː (daru)	*'(door) of the house'*
	house-masc.,sg.,Obl.|Gen-masc.,sg.,Nom.|door-masc.,sg.,Nom.	
(fem.)	gʰara jiː (dariː)	*'(window) of the house'*
	house-masc.,sg.,Obl.|Gen-fem.,sg.,Nom.|window-fem.,sg.,Nom.	

Adjectives agree with the noun they modify in gender, number, and case, marked through the same kind of alternation of the thematic (final) vowel: *nanɖʰoː gʰaru* 'small house, Nom.' *nanɖʰaː gʰara* 'small houses, Nom.', *nandʰiː meza* 'small table, Nom.', *nandʰiyũː mezĩ̃* 'small tables, Nom.'. Pronouns are marked for number, but case is marked only for nominative and oblique (no ablative or vocative pronoun forms). Gender is distinguished only in the nominative, third-person singular forms. The third-person pronouns are also used for demonstratives, for which there are distinct proximal and distal forms: *hiːu* 'this, masc.', *hu* 'that, masc.', *hiːa* 'this, fem.', *hua* 'that, fem.' The personal pronouns shown in Table 3, including dialectal variants, and similar patterns of declension occur with the distinct relative and correlative pronouns (*joː* relative 'who/that-masc.,sg.,Nom.'; *soː* co-relative 'who/that-masc., sg., Nom.'), and the indefinite pronouns (*koː* indefinite 'anyone/someone-masc.,sg.,Nom.'). Of the interrogative pronouns, only 'who' is marked for number and case (e.g. *keru* 'masc.,sg.,Nom.', *kã̄hĩ* 'masc./fem.,sg.Obl.').

TABLE 2

Masculine:	- u *'flower'*	- o: *'boy'*	- u: *'resident'*
Nom.sg.	gulu	tʃʰokiro:	raha:ku:
Nom.pl	gula	tʃʰokira:	raha:ku:
Obl.sg.	gula	tʃʰokire:	rahaku:a, raha:kui
Obl.pl.	gulani	tʃʰokirani	raha:kuani, rahakuni
Abl.sg.	gulã:	tʃʰokirã:	raha:kuã:
Abl.pl.	gulaniã:	tʃʰokiraniã:	raha:kuaniã:
Voc.sg.	gula	tʃʰokira	raha:ku
Voc.pl.	gula:	tʃʰokira:	raha:kua:

Feminine:	- a *'table'*	- i *'eye'*	- a: *'world'*	- i: *'notebook'*
Nom.sg.	meza	akʰi	duniya:	ka:pi:
Nom.pl	mezũ:	akʰi	duniya:ũ:	ka:piũ:
Obl.sg.	meza	akʰiũ:	duniya:	ka:pi:a, ka:pia
Obl.pl.	mezuni	akʰiuni	duniya:uni	ka:piuni
Abl.sg.	mezã:	akʰiã:	duniyã:	ka:piã:
Abl.pl.	mezuniã:	akʰiuniã:	duniya:uniã:	ka:piuniã:
Voc.sg.	meza	akʰi	duniya:	ka:pi:
Voc.pl.	mezũ:	akʰiũ:	duniya:ũ:	ka:piũ:

TABLE 3   Personal Pronouns

		First Person	Second Person	Third Person (Proximal)
Singular	Nom.	a:ũ:, ã: ; mã:, mũ:	tũ:	hi:, hi:u (masc.); hi:, hi:a (fem.)
	Obl.	ã:, mũ:, mũ ; mã:	to:	hina
	Gen.	ã:-jo:, mũhũ-jo:, mũ:-jo:	tũhũ-jo:, tũhĩ-jo:	hina-jo:
Plural	Nom.	asĩ:	tavahĩ:, tavĩ:, tahĩ: ; avahĩ:	hi:, he:
	Obl.	asã:, asã:hĩ, asã:hũ	tavahã:, tahã: ; avahã:, ahã:	hinani
	Gen.	asã:-jo, asã:hĩ-jo	tavhã:- jo ; avhã:-jo	hinani-jo:

Verbs

The verb complex in Sindhi consists of a primary verb, alone or followed by an auxiliary verb. An operator or modal verb element may also occur, placed in between the primary verb (in participle form) and the auxiliary verb. These parts combine in various ways to produce 17 distinct finite verb forms that encode aspect (perfective, imperfective, unspecified), tense (past, present, future), mood (subjunctive, imperative, presumptive, counterfactual), and concordance (gender and number). Aspect is expressed in the choice of the primary verb form, and is marked by suffixation to the verb stem. In most finite verb forms, tense, mood, and concordance features are expressed on the auxiliary verb, through suffixation or auxiliary verb stem allomorphy. There are also six nonfinite verb forms that function as nominal, adjectival, and adverbial participles. Examples of finite and nonfinite verb forms are shown in Table 4. Each finite and nonfinite verb form can undergo further modification, not shown here, to express voice (active/passive) and valence (transitive/causative) distinctions through the use of suffixes that attach directly to the verb stem.

Syntax

The neutral word order in Sindhi is Subject–Object–Verb, although these elements can be permuted in any order, in which case the first element is typically the Topic. Within phrases, word order is fixed, with the head element always at the end. Thus, the determiner and adjective precede the head noun in a noun phrase: *hi:a suhiɳi: tʃo:kiri:* this|beautiful|girl 'this beautiful girl', *asã:ja: ɓa ɓakra:* our|two|goats 'our two goats'. Similarly, the verb appears at the end of the verb phrase, following any (nonsentential) arguments: *ama:-kʰe tʃiʈʰi: likʰi:* mother-to|letter|wrote 'wrote a letter to mother'.

The subject of the sentence appears in the nominative case with two exceptions: "experiencer" subjects of verbs expressing physical sensation, psychological state or kinship are in the dative case, and the ergative subject of a transitive verb in the perfective aspect is in the oblique case. The verb agrees with the subject if it is nominative, and otherwise agrees with a nominative (inanimate) object if present. If there is no nominative subject or object, the verb takes as default the third person, masculine, singular agreement. An indirect object appears with the dative postposition. There is

TABLE 4 Examples of Finite Verb Forms with the Verb Stem *hal-* 'to go'

Contingent future	hale:	'if he goes'
Present unspecified	hale: tʰo:	'he goes'
Contrafactual unspecified	hale: ha:	'had he gone'
Definite future	halando:	'he will go'
Present habitual	halando: a:he:	'he goes'
Past habitual	halando: huyo:	'he used to go'
Presumptive imperfective	halando: hundo:	'he is probably going'
Subjunctive imperfective	halando: huje:	'(perhaps) he goes'
Present continuous	hali: rahyo: a:he:	'he is going'
Past continuous	hali: rahyo: huyo:	'he was going'
Unspecified perfective	halyo:	'he went'
Present perfective	halyo: a:he:	'he has gone'
Past perfective	halyo: huyo:	'he had gone'
Presumptive perfective	halyo: hundo:	'he must have gone'
Subjunctive perfective	halyo: huje:	'he may have gone'
Past iterative	halyo: the:	'he would go', 'he often went', 'he used to go'
Imperative	halu	'Go!' (informal)
	halo:, halije:	'Go!' (polite)

Examples of Nonfinite Verb Forms with Various Stems

Infinitive	halaɳu	'to go'
	ɖa:iɳu	'to sing'
Adjectival unspecified (masc.,sg.)	ma:riɳo:	'about to be struck'
	waʈʰiɳo:	'about to be taken'
Adjectival imperfective (masc.,sg.)	halando:	'going'
	ma:ri:ndo:	'striking'
Adjectival perfective (masc.,sg.)	halyalu, halyo:	'went'
	ma:ryalu, ma:ryo	'struck'
Adverbial imperfective	halande:	'(as he was) going'
	ma:ri:nde:	'(as he was) striking'
Adverbial perfective	hali:, hali: kare:	'(as he) went'
	ma:re:, ma:re: kare:	'(as he) struck'

no distinct accusative case; a direct object appears with the dative postposition for animate objects, and is in the nominative case for inanimate objects.

Examples of verb agreement with nominative case subjects

hu:a hale: tʰi:
they,**fem.**,Nom.| go| Aux.,3pl.**fem.**
'They (fem.) go.'

hu: tʃitʰi: paɽʰanda:
they,**masc**.,Nom.| letter,fem.,sg.,.Nom.| will read,3pl.**masc.**
'They (masc.) will read the letter.'

Examples of verb agreement in sentences with oblique case ergative and experiencer subjects

hunani tʃitʰi: paɽʰi: a:he:
they,masc.,Obl.| letter,**fem.sg.**,Nom.| read,perf.,**3sg.,fem.**| Aux,**3sg**.
'They (masc.) have read the letter (fem.).'

tavahã: kʰe: hunani sã: ʋaɲaɳo pavando
you,pl.masc.,Obl.| Dat.| them,pl.,Obl.| with| go,Inf.,Obl.| must,**3sg,masc**.
'You (pl.) must go with them.'

Historical Development and Sociopolitical Factors

Sindh is home to the ancient Indus Valley civilization of Mohen-jo-daro, but a historical link between Sindhi and the language of Mohen-jo-daro has not yet been firmly established. Sindh has been subject to foreign rule at many times in its history, and there is evidence of language contact in numerous borrowings from Persian, Arabic, and English. The formation of Pakistan in 1947 had a tremendous impact on Sindhi, as millions of mostly Urdu-speaking immigrants came to Sindh from India. The immigrants were under no pressure to learn Sindhi, and enjoyed the advantage of speaking Urdu, the language chosen as the national language of Pakistan. On the other hand, Sindhi speakers, especially those in urban areas, were compelled to learn Urdu and, in recent years, English. Bughio's (2001) sociolinguistic study reports census data that paint a bleak picture of the diminishing status of Sindhi. Census data from 1941 show that 82% of the Sindh population claimed Sindhi as their mother tongue, while in 1981 this number fell to 52%. Sindh has a large rural population, with 57% of the total population, and in rural areas Sindhi is the mother tongue for 78% of the population, compared to only 18% in

urban areas. The rural and urban populations of Sindh also differ in literacy, with 51% literacy rates in urban areas compared to only 16% in rural areas of which females constitute only 5%. The decline in Sindhi literacy has implications for the survival of the rich body of Sindhi literature, dating from the sixteenth century (Ajwani 1970).

References

Ajwani, Lalsingh H. 1970. History of Sindhi literature. New Delhi: Sahitya Akademi (reprinted 1994).

Bughio, Qasim. 2001. A comparative sociolinguistic study of rural and urban Sindhi. Muenchen, Germany: Lincom Europa.

Grierson, Sir George A. 1919. Linguistic Survey of India, Vol. VII, Part I: Indo-Aryan family: northwestern group, specimens of Sindhi and Lahnda. Calcutta: Superintendent Government Printing.

Khubchandani, Lachman M. 1961. The phonology and morphophonemics of Sindhi. Unpublished M.A. thesis University of Pennsylvania.

Masica, Colin P. 1991. The Indo-Aryan languages. Cambridge: Cambridge University Press.

Nihalani, Paroo. 1986. Phonetic implementation of implosives. Language and Speech 29. 253–62.

Nihalani, Paroo. 1995. Sindhi. Journal of the IPA 25(2). 95–98.

Rohra, Satish K. 1971. Sindhi, Kacchi, and emigrant Sindhi. Indian Linguistics 32(2). 123–31.

Stack, Captain George. 1849. Dictionary, English and Sindhi. Bombay: American Mission Press.

Trumpp, Ernest. 1872. Grammar of the Sindhi language. London: Trübner and Co. and Leipzig: FA Brockhaus.

Yegorova, R.P. 1971. The Sindhi language. Moscow: Naukua Publishing House.

JENNIFER COLE

Sino-Tibetan Languages

The Sino-Tibetan languages are spoken by (a) approximately 1.2 billion speakers of Chinese and (b) over 30 million speakers of Tibeto-Burman. The language family extends over a vast linguistic area stretching some 3,500 miles across much of eastern Asia, the Tibetan plateau, and southeast Asia — north- and eastward from the northeastern corner of China along the eastern coasts, southward along the Irrawaddy and Salween Rivers to the Bay of Bengal, and westward along the Himalayas, extending further down to Nepal and India. Besides, Chinese is spoken by several million overseas Chinese in communities across the world. Sino-Tibetan languages comprise the world's second largest language family after Indo-European. As the preceding figures show, Chinese (Sinitic) accounts for most of the family's total speakers.

Despite recent progress, opinions are still sharply divided as to the membership of Sino-Tibetan. Many Chinese scholars hold that Hmong-Mien (previously called Miao-Yao) and Tai-Kadai families should be included, a quarto-partite scheme proposed by Fang-Kuei Li (1973 [1937]). This hypothesis is generally out of favor with other Sino-Tibetanists, who, influenced by Paul Benedict, believe that the large-scale resemblances of structure and vocabulary among the languages in question are due to diffusion rather than genetic inheritance. Some linguists connect Sino-Tibetan with Austronesian, e.g. August Conrady (1916), Laurent Sagart (1993). Others link Sino-Tibetan with Yeniseian and North Caucasus languages in Northern Siberia, e.g. Sergei Starostin (1984). These speculations seem to be based on incomplete or inappropriate evidence, which cannot be substantiated at the present state of knowledge.

The internal composition of Sino-Tibetan, as narrowly defined, appears to be less problematic, despite lack of general consensus. The two major divisions of Sino-Tibetan — Chinese and Tibeto-Burman — can be divided into seven subgroups each. Figure 1 illustrates the genetic schema.

The exact number of Tibeto-Burman languages or dialects is not known, partly because access to some of them by scholars from outside is difficult, and partly because the names and designations of a number of languages are quite confusing. Thus, some languages listed under different names may turn out to be the same language, while others that are lumped together under one name may in fact be different languages altogether. Besides, a number of languages share the same name with non-Tibeto-Burman languages. For example, the Nung language of Tibeto-Burman must be distinguished from the Nung language of the Tai-Kadai family in Guangxi and North Vietnam. Similarly, the Kiao/Kiu language of Tibeto-Burman in China is not the same as that of the Dravidian language in India. Ethnologue lists some 300 languages, a significant number of which are found across national boundaries. Many still await description. A list of Tibeto-Burman languages and dialects with subgroup affiliation is given in James Matisoff (2003:695–704).
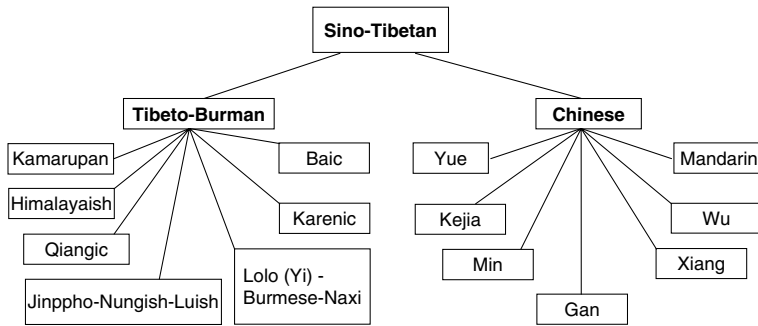
Figure 1. Proposed Sino-Tibetan family tree.

In terms of diversity, India has the largest number of languages listed (137), followed by Myanmar (around 80) and China (around 60). Of these, several languages number over a million (Burmese, some 25,000,000; Tibetan, Yi [previously called Lolo], Kham, Tujia, each with over a million). A small number of languages have probably less than a hundred speakers or semispeakers (Xiandao or Samtao, Lui, Samei).

Sinitic languages appear to be more homogeneous in mutual intelligibility than their Tibeto-Burman counterparts. This is partly because they use a common writing system as a lingua franca. Also, it is customary to talk about Chinese as having seven main dialect groups (q.v.), each with several subdialects, but the actual number of languages or dialects could be much higher. One might even count them in the hundreds, as mutual intelligibility can be low (e.g. Mandarin vs. Cantonese; Wu vs. Kejia) and they are quite diversified. Indeed, dialect differences can be so drastic that, in some cases, varieties could reasonably be considered distantly related languages. This diversity probably speaks for dual (Altaic and Austronesian) or even multiple origin of the Chinese language, as suggested by a number of scholars in the field.

Several writing systems are used for Sino-Tibetan languages. Separate traditions characterize Chinese, Tibetan, and Burmese. The Chinese writing system goes back at least 3,000 years. It contains pictographic elements that are preserved in the oracle and animal bone inscriptions and in bronze inscriptions. The Tibetan writing system was derived from the seventh-century Gupta script in a form representing the northwestern variety of the ancient Nam and Zhang-zhung languages. The Burmese writing system dates back to at least the eleventh century. It was based on Old Mon and Pyu, an extinct Sino-Tibetan language found to have existed in Burma since AD 500. The Tomba script of Naxi (Moso) was developed around the twelfth–thirteenth century; this pictographic script was used mainly for religious purposes. Xixia (His-Hsia or Tangut), a Tibeto-Burman language developed in the eleventh to thirteenth century in northwestern China, now extinct, was based on the Chinese model.

The assumed ancestral language Proto-Sino-Tibetan is thought to have split some 6,000 – 7,000 years ago, with the homeland being somewhere on the Himalayan plateau. Proto-Tibeto-Burman is estimated to have a time depth of around 6,000 years (Matisoff 1991). After their split with Tibeto-Burman, Han Chinese at first flourished in the Middle Kingdom (roughly present Shaanxi, Henan, Hebei, Shandong and vicinities) before they established a strong presence south of the Yangtze River around the beginning of the current era (221 BC) (Chang 1999). In the course of intermingling with the coterritorial non-Chinese populations, Chinese speakers have absorbed elements of material cultures and vocabulary from their neighbors. An example is the Yue and Min dialects in south China which possess features that are found to have been borrowed from Kam-Tai, indicating intense contact with the latter. In religion and philosophy, the influence of Indian culture is prevalent, particularly in Tibeto-Burman.

The Sino-Tibetan languages share a set of typological features. The vast majority of lexical roots in Sino-Tibetan languages are monosyllabic. Most Sino-Tibetan languages use intonational pitch (tone) to distinguish word meanings, with the exception of the Bodo-Garo languages of Assam in India and a few Tibetan dialects such as Purik and Ambo, which lack tones. The overwhelming majority of Sino-Tibetan languages possess noun classifiers. Derivational and morphological affixes can be shown to have existed at one time in Sino-Tibetan, for which reconstructions have been proposed at the Proto-Sino-Tibetan level: *s-* (causative), *-s* (nominalizing), *-t* (transitivizing). Morphological processes are characteristically represented in initial consonant alternation and vowel alternation to produce 'word families', that is, variant forms that are partially related to each other phonologically and semantically, comparable to 'doublets' in Indo-European.

However, the Sino-Tibetan languages are quite different in word order. Sinitic languages have the S(ubject)–V(erb)–O(bject) and modifier–modified word order, Tibeto-Burman are basically SOV and

modified–modifier, except for the Karenic languages of Burma and Thailand, the Bai language of Yunnan and Tujia of Hunan and Hubei, which have the SVO order like Chinese.

The idea of Sino-Tibetan as a language family is based on the assumption that Sino-Tibetan languages are all related and are derived from a common source. Since the late nineteenth century, particularly the early twentieth century, scholars have made various attempts to verify the Sino-Tibetan hypothesis. This typically involves searching for cognates (words derived from the same historical source) by comparing reconstructed forms between Chinese and Written Tibetan. Among the important successes were Robert Shafer's multivolume *Introduction to Sino-Tibetan* (1966–1974), as well as Benedict's (1972) *Sino-Tibetan: A conspectus*, where a sizeable number of basic words are shown to be related, such as the following:

English	Chinese	Tibetan
sun	*njit*	*nyi-ma*
moon	*nguat*	*s-ng^w(y)at*
eye	*myak*	*mig*
fire	*h^{w}ər*	*me*
fish	*ngjag*	*nya*
I	*ngag*	*nga*

Subsequent efforts include Coblin (1986), where some 500 cognate sets are proposed. Austin Hale (1982) offers a critique of the research situation up to the early 1980s. A state-of-the-art survey is provided in Thurgood and LaPolla (2003).

On the Chinese front, the great Swedish sinologist Bernhard Karlgren conducted systematic studies of Chinese dialects and the phonetic series from historical and philological sources. He also considered Chinese loanwords in neighboring languages such as Japanese. On the basis of this combined evidence, he was able to propose several thousand reconstructed forms for Old Chinese. These serve as the basis for subsequent enquiries. Further work along these lines has helped to refine the reconstructions (William Baxter 1992). Although much remains to be done, a basic phonological system of Proto Chinese can be established on the basis of the available materials.

Similar achievements characterize Tibeto-Burman studies. The 1980s and the early 1990s witnessed an influx of new data, with detailed descriptions of a significant number of individual languages. Particularly worth noting are two major comparative dictionaries by Chinese researchers (Sun et al [eds.] 1991; Dai, Huang et al. [eds.] 1992), which contain valuable information on some 50 Tibeto-Burman languages in China. Bradley (1997) reports on research in Tibeto-Burman languages of the Himalaya region. In America, a large

research collaboration, the Sino-Tibetan Etymological Dictionary and Thesaurus (STEDT) Project, was initiated in Berkeley in 1987 under the directorship of James Matisoff, culminating in an 800-page monograph, *Handbook of Proto Tibeto-Burman* (Matisoff 2003). This impressive volume contains several hundred cognate sets for Tibeto-Burman. It offers many insightful observations and discussions of morphological and derivational processes, which firmly establish Tibeto-Burman as a major branch of Sino-Tibetan in its own right. However, due to insufficient data, the handbook is able to provide only partial reconstructions of Proto-Tibeto-Burman forms. Problems involving reconstructed tones for Tibeto-Burman have yet to be resolved. A number of proposed protoforms still cannot be verified due to incomplete data. Several etyma are set up provisionally without having established step-by-step sound correspondences in advance. These inefficiencies can only be overcome when solid work on the reconstructions of individual subgroups is carried out. In addition, the issue of loan contact is becoming more acute as it is often exceedingly difficult to separate early loans from inherited cognates. The position of Tujia and Bai within Tibeto-Burman needs to be further examined.

Sino-Tibetan represents one of the oldest and most diverse civilizations and cultures in the world. Although remarkable progress has been made, the field is still in its infancy. Thus, most proposals for external affiliations and internal subgroupings must be considered working hypotheses rather than firm conclusions. Deeper investigations, in collaboration with allied fields of archeology, history, anthropology, and human genetics, will allow verification of these hypotheses, and will enhance our understanding of the complexities of cultural history in this region.

References

Baxter, William. 1992. A handbook of old Chinese phonology. Trends in Linguistics, Studies and Monographs No. 64. Berlin/New York: Mouton de Gruyter.

Benedict, Paul K. 1972. Sino-Tibetan: a conspectus (Contributing editor: James Matisoff) (Princeton–Cambridge Studies in Chinese linguistics, Vol. 2.). Cambridge and New York: Cambridge University Press.

Bradley, David. (ed.) 1997. Tibeto-Burman languages and classification. Tibeto-Burman languages of the Himalayas, 1–72 (Papers in Southeast Asian Linguistics, No.14). Pacific Linguistics, Series A, No. 86. Canberra: Department of Linguistics, Research School of Pacific and Asian Studies, Australian National University.

Chang, Kwang-Chih. 1999. China on the eve of the historical period. The Cambridge history of ancient China: from the origins of civilizations to 221 B.C., ed. by Michael Loewe and Edward L. Shaughnessy, 37–73. Cambridge: Cambridge University Press.

Coblin, W. South. 1986. A sinologist's handlist of Sino-Tibetan lexical comparisons, Monumenta Serica Monograph Series xviii. Nettetal: Steyler Verlag.

Conrady, August. 1916. Eine merkwürdige Beziehung zwischen den austrischen und den indochinesischen Sprachen. Aufsätze zur Kultur und Sprachgeschichte, vornehmlich des Orient, Ernst Kuhn zum, vol 70, 475–504. Breslao Marcus.

Hale, Austin. 1982. Research on Tibeto-Burman: languages. Trends in Linguistics, State-of-the-art Reports, 14. Berlin: Mouton.

Karlgren, Bernhard. 1957. Grammata Serica Recensa. BMFEA 29. 1–332.

Li, Fang-Kuei. 1973. Languages and dialects of China. Journal of Chinese Linguistics 1. 1–13.

Matisoff, James A. 1991. Sino-Tibetan linguistics: present states and future prospects. Annual Review of Anthropology 20. 469–504.

Matisoff, James A. 2003. Handbook of Proto Tibeto-Burman: system and philosophy of Sino-Tibetan reconstruction. Berkeley/Los Angeles/London: University of California Press.

Sagart, Laurent. 1993. Chinese and Austronesian: evidence for a genetic relationship. Journal of Chinese Linguistics 21. 1–65.

Shafer, Robert. 1966–1974. Introduction to Sino-Tibetan, 5 parts. Wiesbaden: Harrassowitz.

Starostin, Sergei A. 1984. On the hypothesis of a genetic connection between the Sino-Tibetan languages and Yeniseian and North Caucasian languages. In: Linguisticheskaja rekonstrukcija i drevnjaja istorija Vostoka, Vol 4, 19–38. Moskow: Nauka.

Thurgood, Graham. 2003. Sino-Tibetan subgroupings. Sino-Tibetan languages, ed. by Graham Thurgood and Randy LaPolla, 3–21. London and New York: Routledge.

YONGXIAN LUO

Slobin, Dan Isaac

Dan Isaac Slobin is a prolific data-gatherer and avid investigator of child language acquisition. Growing up in the polyglot, immigrant world of Russian Jewish immigrants in the US Midwest, and hearing English, Yiddish, Russian, and Hebrew around him, Dan learned the languages, music, and literature of these cultures. With a father who was a high school teacher of history who loved to travel, Dan and his family lived in Europe and traveled all over the United States and Mexico. As a result, Dan Slobin became hooked on language, languages, and cultures. Philosophically, he was fascinated by the question of how a child becomes a member of a culture, focusing especially on problems of unconscious learning, underlying structure, and mental organization. When he was entering academia in the early 1960s, language was the most obvious and accessible domain in which to look for underlying patterns and their development from a psychological point of view.

Being at the right place and in the right time to take part in the emergence of cognitive science from the world of behaviorism, Dan Slobin's research career has attempted to synthesize the perspectives and insights of the diverse teachers and colleagues from whom he learnt at Harvard and MIT: George Miller, Jerome Bruner, Roger Brown, Eric Lenneberg, Noam Chomsky, and Roman Jakobson.

Dan Slobin was a graduate student at Harvard (1960–1963) when the cognitive revolution was in its infancy. Under the influence of the first phase of Chomsky's transformational grammar, Dan's dissertation, *Grammatical transformations in childhood and adulthood* (Harvard University, 1963), used reaction time to study syntactic processing in both adults and children. His dissertation, in which he found an interaction between semantics and syntax, was one of the first counterthrusts to purely syntactic processing. The dissertation was also significant as it incorporated the diverse interests of his three mentors: George Miller (experimental psychology), Jerome Bruner (cognitive psychology), and Roger Brown (child language development). Under George Miller, Dan Slobin looked for influences of grammatical transformations of reaction time. From Jerome Bruner, Dan considered whether people would pay attention to meaning, as well as form. Additionally, under the influence of Roger Brown, he added a developmental dimension to his dissertation project by including children as well as adults. This early interdisciplinary perspective would set the tone for his research career.

After leaving Harvard with a Ph.D. in Social Relations, Dan Slobin took his first job in the Psychology Department at the University of California at Berkeley, where he has remained throughout his illustrious career. At Berkeley, he was influenced by a group of colleagues who added social, interpersonal, and cultural dimensions to his work, which has grown to include crosslinguistic, cognitive, sociolinguistic, and psycholinguistic perspectives. In California, he worked with Susan Ervin-Tripp, John Gumperz, John Searle, Irving Goffman, Dell Hymes, and Charles Ferguson.

Dan Slobin is credited with focusing the field on the importance of addressing crosslinguistic data when exploring universals in child language acquisition by proposing that different types of languages pose different types of acquisition problems. His research over the past 35 years has been especially concerned with processes of language acquisition and their relations to cognition and thinking on the one hand, and to culture and communication on the other. Throughout his career, Dan has used a crosslinguistic research design in an attempt to identify universal developmental patterns, and has formulated a general model of 'operating principles' for the development of grammar. His multivolume *Crosslinguistic study of language acquisition* summarizes this work and provides a comprehensive contribution to the field of child language acquisition.

Since the late1990s, Dan Slobin has focused on the effect of different languages on psycholinguistic processes, including development and adult behavior, in order to account for specific semantic notions that are encoded in the morphology and syntax of the world's languages (e.g. the origins of grammar). Slobin's most recent work is on early child language development — in the age range from one and a half to five — in a crosslinguistic framework. His interest is in developing relations between language and thought, looking for universal and language-specific patterns. The current focus in the Child Language Research Laboratory is deaf children's acquisition of sign language as a first language, observing a American Sign Language and Sign Language of the Netherlands (SLN), comparing families in which parents are deaf and use a native sign language with the children, and families in which parents are hearing, and are acquiring sign language as a second language.

Thus, in his most recent work, Dan Slobin is again attempting to synthesize research from historical language change and grammaticization, linguistic typology, and universals, pidgin and creole linguistics, sign language, and the study of various discourse forms, especially narrative (both oral and written). This research incorporates experimental and naturalistic methods, as well as computer-aided text analysis.

Biography

Dan Isaac Slobin was born in Detroit, Michigan on May 7, 1939. He received his B.A. in psychology (1960) from the University of Michigan, Phi Beta Kappa, M.A. (1962), and Ph.D. (1964) (Harvard University) for his dissertation on grammatical transformations in childhood and adulthood. He was mentored by George Miller, Jerome Bruner, and Roger Brown. He has held positions in the Department of Psychology at the University of California (Berkeley) since 1964 (Professor since 1972), and also Research Psychologist, Institute of Cognitive Studies and Institute of Human Development. His other positions include: Visiting Professor, Stanford University (1998–1999); Visiting Professor, Summer Linguistics Institute, University of Goettingen, Germany (1992); Visiting Professor, University of Santiago de Compostela, Spain (1989); and Visiting Professor, Summer Linguistics Institute of the Linguistic Society of America, University of New Mexico, Albuquerque (1980, 1995). His research positions include: Institute of Human Development, University of California at Berkeley (1988); Institute of Cognitive Studies, University of California at Berkeley (1964): Visiting Scholar, Laboratoire de Psychologie, CNRS, Paris, France 1964–1985; Visiting Scholar, Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands (1981, 1984–1985, 1992–1999); Visiting Scholar, Department of Linguistics, Tel-Aviv University, Israel (1983–1985); and Visiting Scholar, Department of Psychology, Bogazici University, Istanbul, Turkey (1969–1970, 1972–1973). He has held numerous administrative and editorial positions, and has received awards and grants including the Harvard Foundation for Advanced Study Prize Fellowship (1960–1961); Guggenheim Fellow (1984–1985); and The New York Academy of Sciences Award in the Behavioral Sciences. He has also been a recipient of grants from The Ford Foundation, The W.T. Grant Foundation, National Science Foundation, National Institutes of Mental Health, American Council of Learned Societies, Social Science Research Council, The Sloan Foundation, Israel–United States Binational Science Foundation, Institute of Turkish Studies, American Psychological Association, International Research and Exchange Board (IREX), and the Mellon Foundation.

References

Berman, R.A., and Slobin, D.I. 1994. Relating events in narrative: a crosslinguistic developmental study. Hillsdale, NJ: Lawrence Erlbaum Associates.

Johnston, J.R., and Slobin, D.I. 1979. The development of locative expressions in English, Italian, Serbo-Croatian and Turkish. Journal of Child Language 6. 531–47.

Slobin, Dan Isaac. 1973. Cognitive prerequisites for the development of grammar. Studies of child language development, ed. by C.A. Ferguson and D.I. Slobin, 175–210. New York: Holt, Rinehart & Winston.

———. 1977. Language change in childhood and in history. Language learning and thought, ed. by J. Macnamara. New York: Academic Press.

———. 1982. Universal and particular in the acquisition of language. Language acquisition: the state of the art, ed. by E. Wanner and L.R. Gleitman, 128–72. Cambridge: Cambridge University Press.

————. (ed.) 1985, 1992, 1997. The crosslinguistic study of language acquisition: (1985) Vol. 1. The data; (1985) Vol. 2. Theoretical issues; (1992) Vol. 3; (1997) Vol. 4; (1997) Vol. 5. Expanding the contexts. Mahwah, NJ: Lawrence Erlbaum Associates.

————. 1985. Crosslinguistic evidence for the language-making capacity. The crosslinguistic study of language acquisition: Vol. 2. Theoretical issues, ed. by D. Slobin, 1157–256. Hillsdale, NJ: Lawrence Erlbaum Associates.

————. 1996. From 'thought and language' to 'thinking for speaking.' Rethinking linguistic relativity, ed. by J.J. Gumperz and S. Levinson, 70–96. Cambridge: Cambridge University Press.

————. 1997. The origins of grammaticizable notions: beyond the individual mind. The crosslinguistic study of language acquisition: Vol. 5. Expanding the contexts, ed. by D.I. Slobin, 265–323. Mahwah, NJ: Lawrence Erlbaum Associates.

Slobin, D.I., and Kuntay, A. 1995. Nouns and verbs in Turkish child directed speech. Proceedings of the 19th annual Boston University Conference on Language Development: Vol. 1, ed. by D. MacLaughlin and S. McEwen, 323–34. Somerville, MA: Cascadilla Press.

KAREN WOODMAN

See Also **Acquisition; Miller, George; Psycholinguistics**

Sociolect

Sociolinguistics deals with linguistic behavior in society and is especially concerned with the language situation in such social institutions as political parties, governmental administration, and other nationwide institutions. Sociolinguistics is also interested in language situations in small groups, families, and all other types of social organizations, addressing the distribution of language varieties in such institutions. It is the scientific study of language as it varies with social differentiation.

The distribution of languages (and dialects) in a society is normally described in terms of the age, sex, education, occupations, and ethnic membership of the speakers who make up the society. Links among various aspects of the sociolinguistic situation derive from sociological constructs that underlie the social group or social institution and serve as the context for the stable arrangement of the various factors in language environments. They are thought to provide an individual with the orientation needed for acceptable social behavior in any particular context.

Language use can be studied in the home, the school, the church, the government, and so on. Its behavioral characteristics are then described and analyzed. Forms of linguistic behavior that depend on social factors are called 'sociolects'. A sociolect is thus a form of language depending on age differentiation, occupational distribution, sex differentiation, regional differentiation, etc.

For example, in parts of Germany (e.g. Swabia), a local dialect is used at home, in contact with a peer group, and in other private conversation, but it is necessary to switch to the standard language in more formal circumstances and in contact with strangers. This constitutes a particular problem for schools, which must cope with the various linguistic forms and must provide opportunities for different parts of the population to use the various linguistic forms in appropriate situations. Such 'diglossic' situations occur in many societies and may be interpreted as a form of social stratification.

Diglossia

'Diglossia' may be defined as the complementary use of two different languages (or varieties) in different situations of language use (contact) in a society. Each language is appropriate in particular situations. The labels 'High' and 'Low' for the varieties is probably more an idealization than a characteristic of any real social environment, but it is reasonable to use them to suggest social relations: that is, the use of certain High and Low forms serves to provide an orientation for speakers, a mechanism for them to confirm their expectations.

Diglossia is related to prestige factors in society: that is, high social prestige may accord with relative competence (or lack of competence) in a High variety, and vice versa. This is a function of the standardization of the High variety. Consequently, mastering the High variety can also serve as a vehicle for social mobility.

Diglossic differentiation, therefore, constitutes a sociological model that parallels other types of social stratification in a society: that is, it reflects general differentiation in a society. A number of sociolinguistic studies show that this diglossic language distribution exists, and examine its effects on the direction and rate of development of the society. However,

while such studies describe the stratification in society, they do not explain it; this may require a purer sociological study.

Stratification

Social stratification in a society is reflected in its linguistic stratification, and sociolinguistics tends to describe the social structure as the consequence of the linguistic stratification (i.e. the existence of sociolects) and its effects on the behavior of individuals. Thus, linguistic behavior becomes a measure of the structure of social institutions such as governmental administration, education, political parties, etc. Furthermore, there appears to be a tendency to measure differences among ethnic, religious, and cultural institutions exclusively along linguistic grounds. The underlying hypothesis, then, seems to be that differences in language behavior are the most important differences in social behavior. However, there is growing evidence that complex social relationships cannot be described solely in terms of their correlation with linguistic behavior; rather, they must be described through models of social function, i.e. sociological models. This is reflected in sociolinguistic surveys concerning maintaining a language within a society vs. shifting from one language variety to another. There is also a notion that a particular language may constitute a hindrance to modernization or development, but this notion must be questioned in light of the revitalization of a language (e.g. Hebrew in Israel) or of the extinction of a functioning language (e.g. Gaelic in Ireland). The fate of any particular language or dialect in a given society is the result of the needs of that society to organize social behavior. There is evidence for such a view both in stable multilingual situations (e.g. Switzerland, where there are four official languages) and in situations where multilingualism gives way to a stable monolingual situation and/or standardization (e.g. Malaysia, where Malay replaces local languages, and in many other developing countries around the world). Thus, the language situation is a result of a variety of social efforts involved in planning a stable society, and language change is the result of the need to change society and to provide what is necessary for the functional organization of that society. And such changes may be results of changes in the rural–urban population distribution, in industrial environments, or in the integration of immigrants into society.

Minority/Majority Group Relations

Another aspect of research into sociolects is the description of the relationship between minority and majority groups in society. Majorities tend to domi-

nate minorities with respect to social, cultural, and linguistic phenomena, as, for example, can be seen with respect to the Bretons, the people of Alsace-Lorraine, and the people of Provence in France, all of whom are rapidly losing their own languages in favor of French. Planning and stabilizing the minority/majority situation cannot subordinate the minority group to majority interests. Language is a very important indicator of interethnic relationships within a social group. Language marks both the social identity of the individual and social status in society. Ethnic groups, like other social groups (e.g. age, sex, and religious groups), rely heavily on language to mark their separation from other groups — their distinctness. The language factor not only involves the recognition of other groups as different and the accompanying development of cultural stereotypes; it also takes in the organization of all of society: thus, language is not to be interpreted as merely the medium of communication; it is in fact the 'symbol of ethnicity' par excellence.

Intergroup relationships become more complex and more important as ethnic groups experience a greater need to assert their distinctiveness within the larger society. This phenomenon is observable in Great Britain with respect to the Welsh and Scots or in Spain with respect to the Catalans and Basques. The notion that equates culture, society, and language is belied in many instances by the efforts of various (minority) language groups to achieve self-representation and distinctness. The best way to achieve a stable situation, in most instances, would simply be to adopt greater tolerance, because pressure on the minority to accept majority linguistic and social norms tends to create resistance on the part of the minority. Therefore, the majority must tolerate the language of the minority as a sociolect within society. It is apparent that the argument for linguistic unity within a society is an ideological argument and that societies are in fact structured organisms accommodating various language groups and sociolects.

Education

A further area of investigation concerns the interconnection between education and sociolinguistics. Research like Basil Bernstein's (1971) generated interest in the systematic investigation of the educational problems of individuals and groups identified as somehow 'disadvantaged'; outside Britain, more general questions are asked. Although Bernstein's 'code' model provided an insight into the relationship between language and social situation, and thus could be interpreted as a model for the sociolectal organization of society, further research has questioned both the theoretical model and the adequacy of the empirical

data. Sociolinguists are increasingly interested in the overall awareness of language function in education, not only in foreign language teaching but also in mother-tongue teaching. Sociolinguistics can help to answer relevant questions regarding language use in other societies, and mother-tongue education can profit from the sociolinguistic differentiation of components of the teaching process including both teaching materials and techniques. Foreign language teaching must be highly aware of the sociolectal system of a society as well as changes in it. Foreign language teaching was long dominated by psychological models of learning, particularly the behaviorist model, but there has been growing interest in the function of social rules in foreign language learning and use. Such areas as code-switching and the study of the forms and functions of culturally dependent language use (e.g. types of address and the differentiation of varieties with respect to politeness) are also receiving greater attention. In turn, these concerns have led to greater interest in contrastive analysis (across language systems, within language systems, and across cultures). It is increasingly evident that success in foreign language teaching is in part a function of the ability to use the target language socio-functionally: mere accurate manipulation of the vocabulary and grammar of the target language is insufficient; it is also necessary to know the sociolectal system of the foreign society.

References

Bernstein, Basil. 1971. Class, codes, and control. London: Routledge and Kegan Paul.

Ferguson, Charles A. 1959. Diglossia. Word 15. 325–40.

Fishman, Joshua A. 1972. The sociology of language: an interdisciplinary social science approach to language in society. Rowley, MA: Newbury House.

Giles, R., R.Y. Bourhis, and D.M. Taylor. 1977. Towards a theory of language in ethnic group relations. Language, ethnicity, and intergroup relations, ed. by H. Giles. London: Academic Press.

Gumperz, John J. 1958. Dialect differences and social stratification in a North American Indian village. American Anthropologist 60. 668–82.

Hartig, Matthias. 1998. Soziolinguistik des Deutschen, 2nd edition. Berlin: Weidler.

Hartig, Matthias, and Ursula Kurz. 1971. Sprache als soziale Kontrolle. Frankfurt: Suhrkamp.

Haugen, Einar. 1968. Language planning in modern Norway. Readings in the sociology of language, ed. by J. Fishman. The Hague: Mouton.

Labov, William. 1966. The social stratification of English in New York City. Washington, DC: Center for Applied Linguistics.

Rubin, Joan. 1968. National bilingualism in Paraguay. The Hague: Mouton.

Shuy, R.W. 1974. Sociolinguistic strategies for teachers in a southern school system. Proceedings of the 3rd International Congress of Applied Linguistics, ed. by A. Verdoodt. Heidelberg: Julius Gross.

Tanner, Nancy. 1967. Speech and society among the Indonesian elite: a case study of a multilingual community. Anthropological Linguistics 9(3). 15–40.

MATTHIAS HARTIG

Sociolinguistics

Contemporary sociolinguistics is a diverse and thriving field of linguistic inquiry. Its major branches (variationist sociolinguistics, the sociology of language, and the ethnography of communication/interactional sociolinguistics) share an interest in the ways in which human societies organize their social lives in and through language. In recent years, the theoretical models have shifted from static ones, which assumed that the language use of a person was a straightforward expression of their social identity (social class, sex, age, ethnicity, nationality, professional status, and so on), to social constructionist ones, which view language use as a practice that creates social identity. Furthermore, the ways in which power relationships in society constrain the linguistic expression of its members and the interpretation of their utterances have become a central concern.

Linguistics emerged as an academic discipline in the nineteenth century. Some of the founders of the discipline were dialectologists, such as Jules Gilliéron in France or Georg Wenker in Germany. These men were interested in describing regional variation in language in terms of pronunciation, word structure, syntax, and vocabulary, and they produced elaborate dialect atlases for the areas they studied. Their concern with the ways the identities of speakers influence the ways they speak foreshadows the central interest of sociolinguistics. However, sociolinguistics in its modern sense is usually dated from the 1950s. Although work in dialectology continued, mainstream linguistics had marginalized interest in social questions of language use by that time. Language as a system and the rules that govern that system had developed as the central research focus of linguistics, and interest in the

speakers had largely fallen by the wayside. Increasing frustration with this view of language, which seemed very remote from real life, led to at least three new developments, which today make up the three main branches of the diverse field of sociolinguistics.

Dialectology had set out to describe speech differences between speakers of different regional backgrounds. It was a central assumption of that work that such differences could best be observed in rural areas and that NORMs (nonmobile, older, rural males) would speak a dialect in its purest form. Speakers who did not fit this bill, such as city dwellers, were thought to vary their speech haphazardly. New Yorkers, for instance, were known to sometimes pronounce the [r] in words such as *card*, and sometimes not. The general assumption was that there was no pattern to this whatsoever, and the phenomenon was termed 'free variation' until William Labov's work showed otherwise. In the 1960s, William Labov got sales clerks in three New York department stores (one up-market, one medium, and one down-market) to pronounce 'fourth floor' for him by asking where certain products were sold. By pretending to be hard of hearing, he then got them to repeat it. In this way, he got a sample of the [r] use of upper-class, middle-class, and working-class-oriented speakers, in two different styles (casual and careful). He found that the variation was not haphazard at all, but that the working class speakers used [r] in relevant words with the lowest frequency and upper-class-oriented speakers with the highest. Furthermore, all groups used more [r]s in careful pronunciation than in casual pronunciation. These findings of patterned social variation have since been confirmed and elaborated in a number of studies of urban speech variation, mainly in the English-speaking world. All these studies collected speech samples, often through interviews, from a random selection of speakers, representative of social class, ethnic group, sex, age, style (formal vs. casual speech), and so on. Whatever the variable that was studied in a specific community, this research showed that certain variables are indicative of the social background of a speaker. Many of these studies, such as Labov's (1972) work in New York or Horvath's (1985) research in Sydney, have become sociolinguistic classics. However, variationist sociolinguistics has also been heavily criticized for lacking a sophisticated social theory and for reifying social identity by saying, for instance, that working-class women speak in such and such a way. Particularly, language and gender researchers have been concerned with the static picture of female vs. male speech that work tended to paint. The early variationist model did not allow for individual agency, contestation, or change. Consequently, the most exciting work in variationist sociolinguistics is currently inspired by work in language and gender.

Eckert (2000) represents a significant step forward by viewing linguistic variation as social practice. In this view, speakers draw on linguistic resources to project certain identities while rejecting others. Linguistic variation is no longer predetermined by a speaker's place in the social hierarchy, but has become a resource they use to stake their claims to a particular place in their communities.

Another strand of sociolinguistics has been inspired by widespread bi- and multilingualism. For a long time, mainstream linguistics took monolingualism as its paradigm case, despite the fact that the majority of the world's population is bi- or multilingual. In the 1950s, Einar Haugen was one of the first to study language choice by bilinguals in any detail. His central questions continue to be central to the sociology of language: why and how do bilinguals choose one language over the other in a particular situation? Why and how do they sometimes abandon one of their languages and shift to another? Why and how do others maintain their languages? It is usually the use of minority languages that is threatened by languages of wider communication. Minority status may be the result of internal colonization (e.g. Welsh in the United Kingdom), external colonization (e.g. Aboriginal languages in Australia), or migration (e.g. Italian in the United States). State policies have a crucial influence on language maintenance or shift. Such policies may range from genocide and violent repression (e.g. the Aboriginal language of Tasmania in colonial Australia; Kurdish in Turkey) to active promotion of bilingualism (e.g. Singapore or contemporary Australia). Many nations have negative attitudes toward bilingualism, and public discourses about the deficiency of bilinguals negatively affect bilingual usage (e.g. discourses about Spanish in the United States).

Paralleling developments in variationist sociolinguistics, in the 1960s and 1970s, sociologists of language sought to develop taxonomies of social situations that would result in specific language choices. The concept of 'domain' was introduced by Joshua Fishman, who showed that Puerto Ricans in New York consistently chose English or Spanish, depending on whether their linguistic activities occurred in the domain of family, friendship, religion, employment, or education. Since then, code-switching, or the mixing of two or more languages, as often occurs in bilingual speech, has been acknowledged more widely. Furthermore, with the effect of social constructionism on sociolinguistics as a whole, bilingualism researchers have also become wary of seeing language choice as the result of a particular social situation that speakers find themselves in. Rather, they have begun to view language choice as a resource to construct speakers' social identities (similar to the ways in

which monolinguals use different styles to project a particular social identity). Youths in contemporary multiethnic cities around the world have been found to develop their emerging hybrid identities through elaborate language choices. The choices speakers have at their disposal are crucially influenced by power relations in society, and many researchers have found it useful to integrate social constructionist approaches with views that conceptualize languages as cultural currencies—symbolic capital. In this view, language choice is governed by the laws of a market economy: an international language such as English is a hard-currency language, and access to this language is highly desirable but also restricted by its symbolic owners (education boards, television networks, and so on) to safeguard the strength of their currency.

An unresolved (and irresolvable) question for the sociology of language is what a language actually is. Most people think that sociolinguists should at least be able to identify a language unambiguously—that they should be able to draw clear boundaries between languages and be able to say how many languages there are in the world today. They cannot. Even a seemingly unambiguous test such as mutual intelligibility fails: speakers of Danish, Norwegian, and Swedish can understand one another, yet they are said to speak three different languages. By contrast, speakers of Mandarin Chinese and Cantonese Chinese cannot understand one another, despite the fact that they supposedly speak a single language: Chinese. Thus, languages are defined politically rather than linguistically. Max Weinreich is often quoted as having said: 'A language is a dialect with an army and a navy'. Indeed, the development of pidgin and creole languages is another good example of this. Sociolinguists have come to recognize pidgin and creole languages as fascinating instances of language birth. However, outsiders and even speakers themselves often deride these languages as 'not really a language at all' or 'broken English' (in the case of an English-based creole). Only when such languages, which typically evolved in the extreme contact situations of colonialism and slavery, become the official language of a state, such as Tok Pisin in Papua New Guinea, do they get linguistic recognition.

The third major branch of contemporary sociolinguistics is closely connected to anthropological linguistics and also overlaps with conversation and discourse analysis. It is the aim of the ethnography of communication to uncover, describe, and compare the speaking practices of specific communities. Speakers' communicative competence is of central importance in this endeavor. In general linguistics, the concept of competence is used to capture speakers' subconscious knowledge of linguistic rules. Dell Hymes observed that grammatical rules are only half the story. In addi-

tion to the ability to produce grammatically acceptable utterances, speakers also need to know when to speak and when to stay silent, or what is appropriate to say in a particular situation. Dell Hymes was the first to introduce the term 'communicative competence' for these culture-specific speaking rules. Parents spend much more time to socialize their children into socially appropriate speaking rules than teaching them explicit grammatical rules (e.g. *Don't speak with your mouth full!*). Similarly, sharing a common grammatical system does not necessarily guarantee smooth communication, as countless anecdotes about American–British misunderstandings demonstrate.

Ethnographic research is carried out through participant-observation (i.e. researchers become participants in the community they wish to observe). The central goal of this data collection method is the collection of speech that occurs naturally in a social context, rather than being elicited for research purposes. The work of Dell Hymes also provides a framework for the comparison of speaking practices across communities: the so-called SPEAKING grid. SPEAKING stands for setting, participants, ends, act sequence, key, instrumentalities, norms, and genres.

Interactional sociolinguistics developed from the ethnography of communication and is closely associated with the pioneering work of John Gumperz. A central idea is the one of 'contextualization', which posits that certain linguistic cues (e.g. prosody) are used to create a social context and a framework for interpretation. Many such cues are taken for granted, and speakers assume them to be linguistic universals, which in fact they are not. If speakers do not share such presuppositions for the framing and interpretation of communicative interactions, misunderstandings and even communication breakdown may result. Consequently, cross-cultural communication has been of central interest to interactional sociolinguists. Communication in institutional settings has been a further central site of inquiry. Such work, whether on communication in the classroom, in the courtroom, in medical settings, or in the workplace, often has practical applications. It seeks to uncover communication patterns that may give some interactants an advantage over others in such encounters.

From its inception, sociolinguistics has been concerned with power relationships and the differential access to material and symbolic resources that these relationships result in. Many applications result from this fundamental concern. In the field of education, dialect and minority speakers have consistently been found to be less successful than their standard-speaking peers. A number of court cases in the United States have established that such a disadvantage constitutes a form of discrimination. An early one of

these, in Ann Arbor, Michigan, had William Labov testify to the structural equality of African-American vernacular English to other languages. Outside education, sociolinguistic expertise finds applications in cross-cultural and workplace communication, forensic contexts, doctor–patient interactions, and language policy and planning.

References

Ammon, Ulrich, Norbert Dittmar, and Klaus J. Mattheier. (eds.) 1987–1988. Sociolinguistics: an international handbook of the science of language and society, 2 vols. Berlin and New York: Mouton de Gruyter.

Chambers, J.K. 2003. Sociolinguistic theory. Oxford and Cambridge, MA: Blackwell, 1st edition, 1995.

Coulmas, Florian (ed.) 1997. The handbook of sociolinguistics. Oxford and Cambridge, MA: Blackwell.

Coupland, Nikolas, and Adam Jaworski. (eds.) 1997. Sociolinguistics: A reader and coursebook. New York: St. Martin's Press.

Eckert, Penelope. 2000. Language variation as social practice: the linguistic construction of identity in Belten High. Oxford and Cambridge, MA: Blackwell.

Fasold, Ralph. 1986. The sociolinguistics of society. Oxford: Basil Blackwell.

———. 1990. The sociolinguistics of language. Oxford: Basil Blackwell.

Holmes, Janet. 2001. An introduction to sociolinguistics. London and New York: Longman, 1st edition, 1992.

Horvath, Barbara. 1985. Variation in Australian English. Cambridge: Cambridge University Press.

Meshtrie, Rajend, Joan Swann, Andrea Deumert, and William L. Leap. 2000. Introducing Sociolinguistics. Edinburgh: Edinburgh University Press.

Romaine, Suzanne. 1994. Language in society: an introduction to sociolinguistics. Oxford: Oxford University Press.

Wardaugh, Ronald. 1986. An introduction to sociolinguistics. Oxford and Cambridge, MA: Blackwell, 3rd edition, 1997.

INGRID PILLER

See also **Fishman, Joshua A.; Gumperz, John Joseph; Hymes, Dell H.; Labov, William**

Soundwave Analysis

Sound is a form of energy. It is generated by disturbance of particles, and it travels in waves at a certain speed, depending on the compressibility and the resistance of the medium (e.g. 344 m/s in air, 1,500 m/s in water). Most sounds are complex by nature, i.e. they consist of more than one tone, each tone with its own frequency, amplitude, and phase. Over the years, several tools have been developed to analyze speech sounds. The first tools were based on analog techniques, but since the 1950s digital techniques have advanced greatly, and today software packages exist that can be used at home. Digital techniques are more accurate, and are repeatable. One of the most important approaches regarding the analysis of complex sounds (like speech) was proposed by Joseph Fourier, a French mathematician in the second half of the nineteenth century. He showed that a complex speech signal like speech can be considered as consisting of a combination of sine waves (also called sinusoids or pure tones). The sine wave is the simplest of all sounds, and it seems appropriate to discuss its properties first.

Sine Waves

Figure 1 illustrates the three main properties of a sine wave: frequency, amplitude, and phase. The frequency of a wave refers to the repetition of a cycle per unit time (a). The period of a wave is the duration of one cycle, and the frequency is the reciprocal of the period, expressed in Hertz (Hz). The lower the repetition rate, the lower the perceived pitch. Amplitude represents the extent of vibration (b) and is correlated with loudness: the lesser the amplitude, the softer the signal. The phase of the signal is the point in time at which the waveform starts (c). Most sound analysis tools do not take phase differences into account, as they hardly affect the identity of the different sounds.

A complex sound consists of different sine waves, each with its own frequency, amplitude, and phase. In speech, there are two kinds of complex signals: periodic and aperiodic ones. In periodic signals, the vibration repeats itself; in aperiodic signals, the vibration has no repeatable pattern. Vowels usually have a periodic structure due to the continuous vibration of the vocal folds during speech production, while consonants exhibit an aperiodic, noise-like, structure.

Oscillogram

One of the oldest sound analysis tools is the oscillograph (±1930). It displays sound as pressure variations over time and allows the determination of relative amplitude, duration, and fundamental frequency. An example of a digital oscillogram of the word 'soldier'
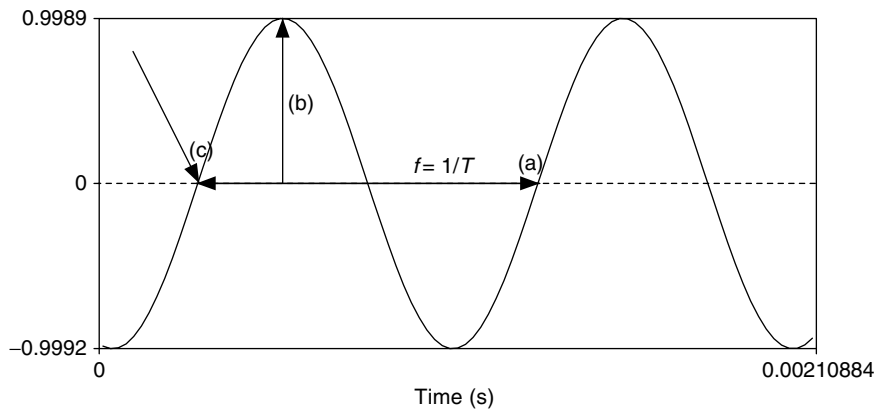
Figure 1

is given in Figure 2A. Note that the vibration of the /s/ is aperiodic, while that of the vowel nucleus of the /o/ is highly periodic. By selecting a portion of the vowel, preferably somewhere in the middle, it is possible to determine its periodicity and hence its fundamental frequency by taking the reciprocal of the period ($f = 1/T$). The fundamental frequency is the lowest frequency and is the physical measure of vocal fold vibration. This physical property is perceived as the pitch of a signal. It is the primary acoustic cue to intonation and stress in speech and is crucial to phoneme identification in tone languages. The determination of $F0$ is usually done for several numbers of periods to minimize measuring artifacts as much as possible. The main drawback of this method is that the measurements, especially determination of duration, strongly depend on the placing of the cursors. For many speech sounds, it is very difficult to determine exactly where a speech segment begins and where it ends. This is illustrated in Figure 2B for the /o/ of /soldier/. The end of the /s/ is easily determined (a), but the end of the /o/ and beginning of /l/ are not marked clearly (b), so that the duration of /o/ cannot be determined reliably. In addition to this, oscillographic analysis does not suffice to describe some of the important differences among vowels. Observation of these differences requires the generation of spectral representations, that is, plots of signal energy vs. frequency (see further).

Root-Mean-Square Amplitude

In order to determine the amplitude of the (speech) signal, both the negative and positive parts of the waveform must be taken into account. This is done by calculating the root-mean-square (RMS) amplitude of a certain number of samples. First, the samples in a waveform window are squared, so that the portions above and below the line are positive. The squared values in the window are then averaged and squared again (to compensate for the initial squaring), after which another set of samples is selected. RMS amplitude

measurements are useful quantities because they relate directly to the power in the signal.

Spectral Analysis

In order to examine the characteristics of speech sounds in more detail, (parts of) the waveform can be submitted to a spectral analysis, also called Fourier analysis. Recall that Fourier showed that periodic waveforms, no matter how complex, can be analyzed as the sum of sine waves varying in frequency, amplitude, and phase. The resulting power spectrum shows the amplitude of each different component (sine wave) as a function of its frequency. These components are also called harmonics in periodic sounds, because the frequencies of the components are multiples of the fundamental frequency or the first harmonic ($F0$). An example of a spectrum of the vowel /i/ is given in Figure 3. The shape of the spectrum is given by the (imaginary) line connecting the amplitudes of the harmonics. Different (speech) sounds have different spectral envelopes, because different vocal tract configurations result in different resonant frequencies. The first component corresponds to the fundamental frequency (and hence to vocal fold vibration). Components yielding the highest amplitude correspond to the resonant frequencies of the sound, here at approximately 300, 2,000, and 3,100 Hz. Spectral analysis has proved to be a very valuable tool in distinguishing the characteristics of the different vowels and consonants (e.g. different frequency regions of /s/ and /f/) (see Figure 3).

Linear Prediction

A different method for estimating formant frequencies is linear predictive coding (LPC). Like the Fourier transform, it relates a representation in time to one in frequency. However, it is not based on spectral analysis. The hypothesis behind LPC is that a sample is partly predictable from its immediate predecessors.

(a)
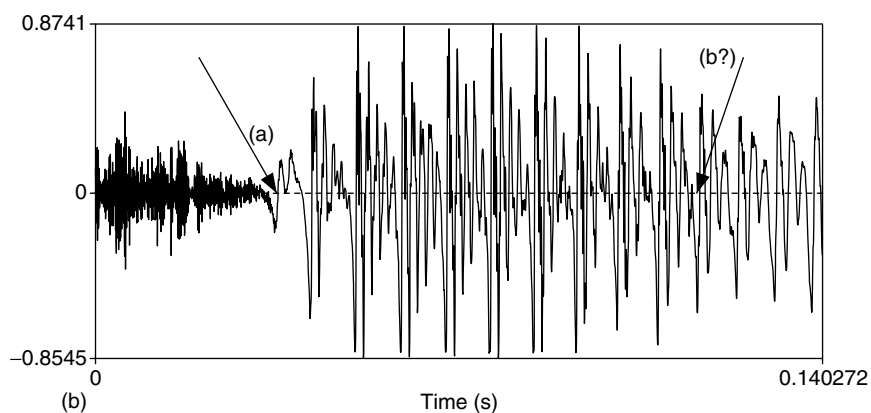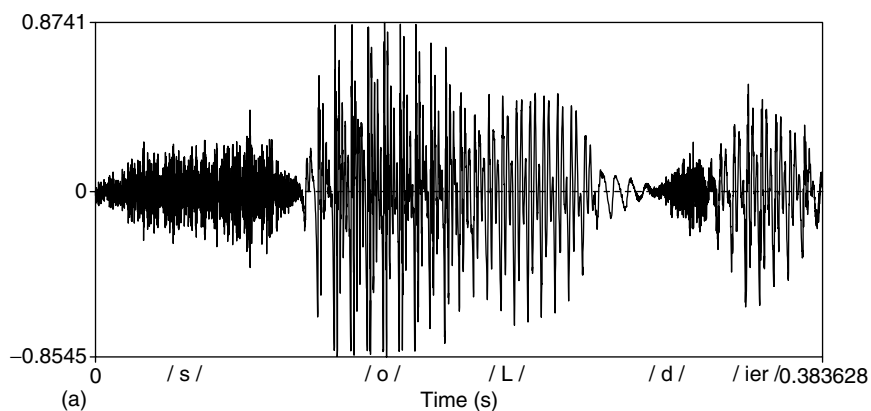
Figure 2

Figure 3
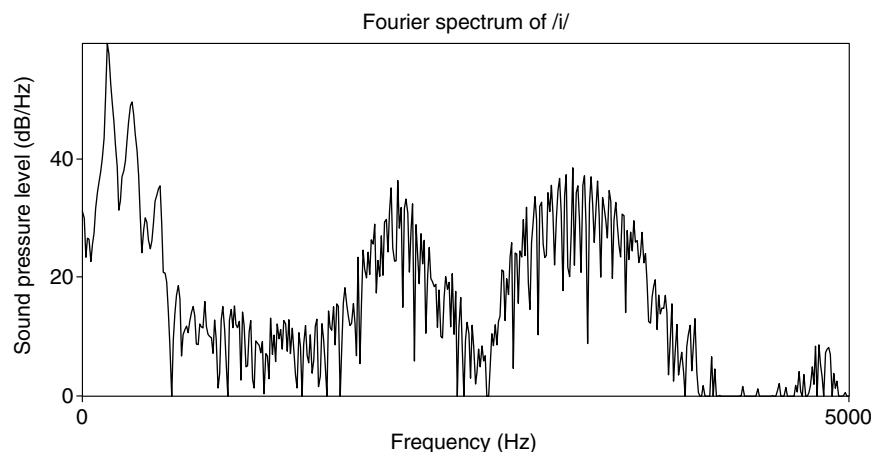
This is true under the assumption that speech does not vary too rapidly from sample to sample (which is not always the case). LPC was originally developed as a method of compressing the signal (e.g. for speech synthesis). The algorithm calculates 'predictor' coefficients that correspond to the vocal tract filter function in a linear way. It does not transmit the individual samples, but determines weights and errors that are updated every 10 or 20 ms, depending on the chosen window size. This is an immense saving compared to the original speech signal. Of course, the prediction is not completely accurate, so the transmitted speech is not perfect. Both the Fourier spectrum and LPC have proved to be good methods for estimating formant frequencies. Note, however, that the two methods yield different types of information. The Fourier spectrum shows the harmonics of the fundamental frequency. Formant frequencies are estimated from the harmonics with the highest amplitude. The LPC spectrum does not show harmonics, but (automatically) yields the formant frequencies and bandwidths from the filter functions. In general, both methods compare well.

There is no 'best' method, and errors occur frequently at high fundamental frequencies. The higher the fundamental frequency, the wider apart the harmonics in the spectrum, and the more difficult it becomes to estimate the resonant frequencies. Moreover, most LPC models only account for the resonances (they are all-pole models), not the anti-resonances (zeroes) in speech sounds. This means that LPC is not suitable for determining formant frequencies in speech sounds that contain antiresonances, such as the nasals and laterals (e.g. /m, n, l/), and some fricatives. Another disadvantage of LPC is that the number of formant frequencies has to be specified before the analysis. For every two coefficients, the algorithm has to find a peak in the spectrum. If the number of anticipated peaks is larger than the actual number of peaks, the algorithm will yield values that do not correspond to true peaks in the spectrum. Similarly, if very few peaks are specified, some of the true peaks will not be indicated either. All in all, it is best to use both methods on the same speech sounds, and to compare results. In this way, errors, such as, for instance, confusing a harmonic with a formant frequency, can be avoided.

Filtering

Filters are commonly used in sound analysis. Filters pass energy at certain frequencies but not at others and therefore allow a selective look at the energy in a certain frequency region. A low-pass filter removes all frequencies above a given cut-off frequency; a high-pass filter removes all frequencies below a given cut-off frequency. A band-pass filter allows frequencies between two cut-off frequencies to pass (our telephone network between 300 and 3,400 Hz !), while a band-stop filter removes the information between two given frequencies. In speech, a preemphasis filter is often applied to boost the high-frequency components in amplitude relative to the low-frequency components. It is actually a high-pass filter with a response that usually increases at 6 dB per octave. A preemphasis filter is necessary, because most of the energy in speech is in the lower frequency range. This energy will tend to dominate the analysis if there is no equalization of energy across frequency. Also, before sampling, the input signal must be low-pass filtered at the value of the maximum frequency of interest (usually 10 kHz when the sampling rate is 20 kHz). This antialiasing filter sharply attenuates frequencies above half the sampling rate to avoid aliasing effects.

Spectrographic Analyses

Spectrographic analysis, an important method for visualizing and examining (speech) sounds, is based on band-pass filtering. This method determines the energy of each frequency component as a function of time and, by doing so, clearly illustrates the variation in the acoustical properties over time. While the Fourier spectrum yields information of the frequency components of a certain speech sound in a rather static way, the spectrogram displays the dynamical nature of a stretch of speech. In the analog spectrograph, which was developed in the 1940s, the speech signal is analyzed by a set of band-pass filters, each of which only responds to the energy within its frequency band. In the digital version, the spectral amplitude results from a series of different Fourier analyses. The number of samples in the analysis window determines the width of the analysis. A short analysis window corresponds to a wide bandwidth, and a long analysis window to a narrow bandwidth. Figures 4A and B illustrate a wideband and narrowband analysis of 'The soldier saluted the flag.' In both plots, time is given on the horizontal axis, frequency on the vertical axis, and intensity by the blackness of the pattern. However, as a result of the different filter settings, there is a trade-off between temporal and spectral information. Wideband spectrograms (Figure 4A) yield high temporal resolution. They display the individual voicing pulses (vertical striations) as well as the resonant frequencies (broad bands of energy) of the speech sounds. Spectral changes over relatively brief intervals of time are clearly visible. Note the high-frequency turbulence of the /s/, the dynamic formant patterns, the silence preceding unvoiced plosives, etc. In a wideband analysis frequency resolution is low, and the fundamental frequency is difficult to determine, especially when it is smeared together with the lowest formant frequency. It is possible to determine the average fundamental frequency by calculating the number of glottal pulses per second (vertically), but this is a time-consuming and not very reliable method. The fundamental frequency can be better determined from a narrowband spectrogram (Figure 4B), where the fundamental frequency and its harmonics are displayed as equally-spaced horizontal lines within the broader formants. A narrowband spectrogram has a high-frequency resolution, more fine-grained information, but low temporal resolution. Usually, the average fundamental frequency is determined by counting the first ten harmonics (see arrow in Figure 4B), and by dividing the corresponding frequency on the y-axis by ten (see Figure 4).

Instead of spectrographic analysis, it is also possible to make use of special algorithms to track the fundamental frequency in the speech signal. However, these programs also make characteristic errors, such as confusing the first formant (resonant) frequency with the fundamental frequency, doubling the fundamental frequency,
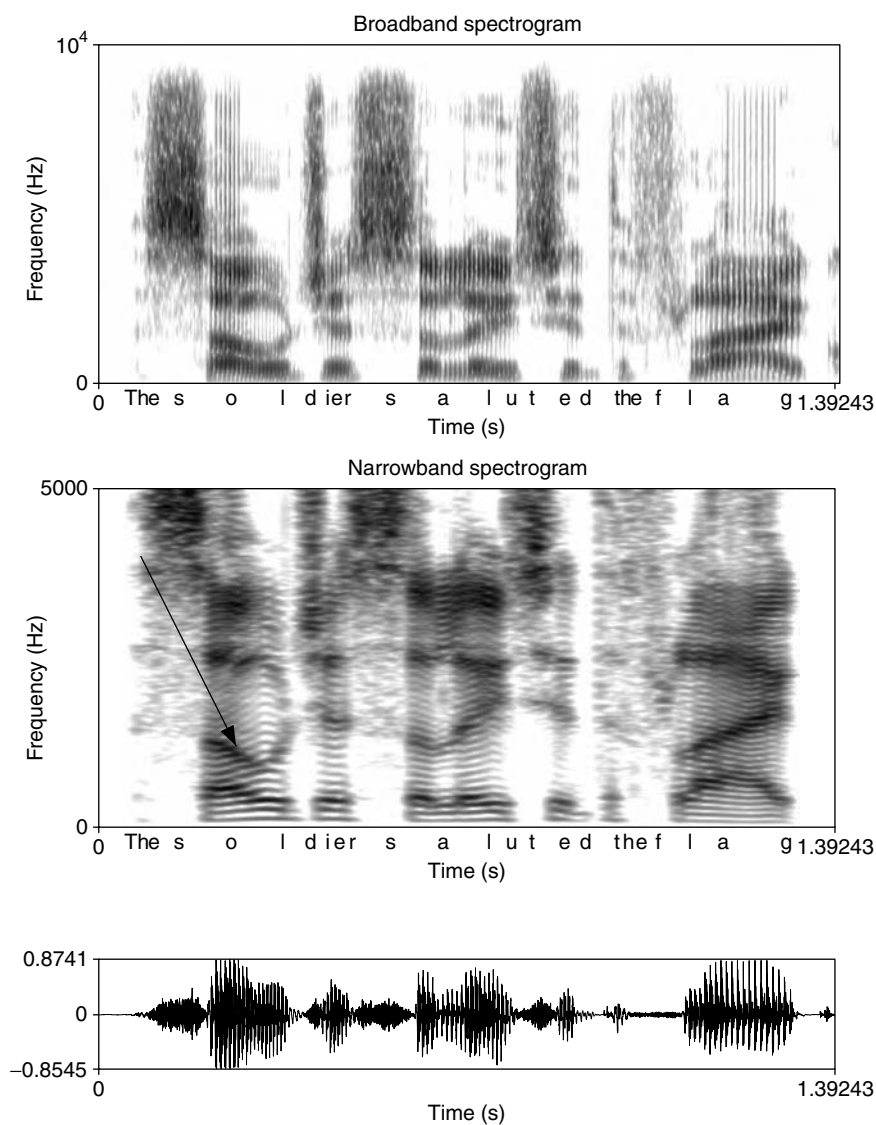
Figure 4

finding a fundamental frequency in unvoiced parts of the signal, or failing to find it in voiced sections.

Digitalization, Sampling, and Quantization

Once (personal) computers became popular, many software programs became readily available for digital (speech) sound analyses. Digital techniques can do what the analog systems did, only more precisely and faster. Moreover, digital signals can be stored forever without quality degradation and the stored signal can be reanalyzed, edited, and displayed as often as necessary. Before analysis, however, the speech signal must be transformed to a discrete representation. This means that the speech signal must be sampled at certain, regular, intervals of time. But what should the sampling rate be? If the rate is too low, much of the information of the original signal will be lost or errors can occur in the analysis. The higher the sampling frequency the better the sound quality, but also the larger the required storage capacity and computer memory. For instance, compact disks are sampled at a rate of 44.1 kHz, meaning that the signal is sampled 44,100 times per second. It is not necessary to sample speech at such a high rate, because the main acoustical properties range between 0 and 10,000 Hz. The Nyquist theorem states that the sampling rate should at least be twice the maximum frequency of interest; hence, a sampling rate of 20,000 Hz is appropriate. Together with sampling (discretization as a function of time) the signal must also be quantized (discretization as a function of level). Quantization is expressed in bits: an 8-bit signal has 256 different levels (2^8), and a 15-bit one has 32,768 different levels (2^{15}). Although the increase in the number of quantization levels results

in a better sound quality (due to the fact that there is a more accurate resemblance between the discretized and original signal) it also implies more memory to store the data. For the analysis of speech sound, 12-bit quantization suffices. However, for the analysis of music, finer gradations of amplitude are required, since the dynamic range, i.e. the difference between the quietest and loudest sound of interest, is greater.

Autocorrelation

Instead of determining the fundamental frequency ($F0$) from the time signal (Figure 2) or from a (narrowband) spectrogram (Figure 4B), it can also be detected automatically by a computer algorithm. Over the years, many computer algorithms have been developed, none of which yield a 100% accurate result (one should always check!). This is because of the nonstationary nature of speech, irregularities in vocal cord vibration, the wide range of possible $F0$ values, and degraded speech in noisy environments. In general, the autocorrelation algorithm yields a higher accuracy than the time-domain techniques, but also requires a higher computation load. In the autocorrelation method, the average correlation of part of the speech signal is computed, usually a Hanning window of a constant length, and a delayed copy of itself (= auto). The 'lag' with the highest correlation is then taken as the pitch of the period. The correlation is highest when the delay is close to one pitch period. Pitch extraction with the autocorrelation algorithm works well, because the spectro-temporal changes in voiced segments are usually relatively slow compared to those in the more dynamic speech segments (e.g. voiceless plosives). However, the algorithm also has its drawbacks. A common error produced by the autocorrelation method is pitch doubling, i.e. when a subcomponent of the pitch period is mistakenly identified as the pitch period. Moreover, formant frequency information can affect pitch extraction, to the extent that the output is not the glottal period, but the glottal period plus the period of the first or second formant. Autocorrelation is one of the more reliable methods of determining $F0$.

Cepstral Analysis

Another fairly accurate pitch extraction method is cepstral analysis. This method makes use of the periodic structure of the harmonics in a Fourier spectrum. To determine the interval between subsequent harmonics, a Fourier transform of the log power Fourier spectrum of the original speech signal is taken. The result is a spike at a component that corresponds to the fundamental frequency. Because of the Fourier transformation on a Fourier transformation, the frequency axis of the spectrum is transformed back to a time axis. Note that 'cepstrum' is actually 'spectrum' with the first syllable reversed. Similarly, the 'quefrency' peak (= 'inverse of' frequency) represents the fundamental period.

References

Fant, G. 1973. Speech sounds and features. Cambridge, MA: MIT Press.
Flanagan, J. 1972. Speech analysis, synthesis and perception. Berlin, Heidelberg, New York: Springer-Verlag.
Furui, S., and M.M. Sondhi (eds.) 1992. Advances in speech signal processing. New York: Marcel Dekker.
Johnson, K. 1997. Acoustic and auditory phonetics, 169pp. Oxford: Blackwell Publishers.
Kent, R.D., and Ch. Read. 1992. The acoustic analysis of speech, 238pp. San Diego, CA: Singular Publishing Group, Inc.
Ladefoged, P. 1994. Elements of acoustic phonetics. Chicago: University of Chicago Press.
Pickett, J.M. 1999. The acoustics of speech communication: fundamental, speech perception theory and technology, 404pp. Newton, MA: Allyn & Bacon.

ASTRID VAN WIERINGEN

South Africa

South Africa, nestled in the southern tip of the African continent, covers 1,211,037 square kilometers. The population of 45.98 million (mid-2000 estimate) includes approximately 26.98 million Africans, 5.99 million Whites, 3.49 million 'Coloreds', and 1.03 million Asians. The African population is made up of Zulu (18.8%), Xhosa (18.3%), Tswana (7.2%), Southern Sotho (6.9%), Northern Sotho (9.8%), Shangaan (Tsonga) (4.2%), Swazi (2.6%), Ndebele (1.5%), and Venda (1.7%). The white population is two-thirds Afrikaans-speaking (a form of Dutch) and one-third English-speaking. The 'Coloreds' are of mixed race, and the Asians are chiefly Indians. The linguistic diversity and complex

history of South Africa poses interesting issues for language students.

South Africa contains two of the four language families (phyla) in Africa: Khoisan and Niger-Kordofanian languages. The Bantu (or Sintu) languages are spoken in South Africa, which are the most widespread and best-known subgroup of Niger-Congo, a subfamily of Niger-Kordofanian family. Sintu reflects the Nguni prefix isi-, denoting language and culture. Sintu languages in South Africa include Nguni languages (Ndebele, Swazi, Xhosa, and Zulu), Sotho languages (North Sotho, South Sotho, and Tswana), Tsonga, and Venda. The Nguni, whose descendants now comprise most of the country's black majority, were established in South Africa by the third century AD. The Nguni languages in South Africa now comprise Xhosa, Zulu, Swazi, Ndebele, Ponso, and Tembu. They are closely related and spoken by most African people in South Africa.

Greenberg (1950, 1963) further developed and popularized the concept of Khoisan as a language phylum. The most prominent feature of the Khoisan languages is the sound system. It contains a large number of ingressive sounds known as clicks, which usually occur word-initially, and are often believed to be remnants of the most ancient sound inventory of human language (Vossen 2000). Also, most Khoisan languages make use of various levels of tone, both lexically and grammatically, and tone-bearing units are syllabic segments such as vowels and nasal consonants. Vowels can occur as oral, nasal, pharyngealized, glottalized, breathy vowels, or combinations thereof.

Several millennia ago, the Khoisan languages predominated all across southern Africa (including most of modern-day South Africa), from the Cape of Good Hope to Somalia and Kenya (Guldemann and Vossen 2000). The Khoisan peoples—Bushmen, Cape Khoikhoi (the so-called Hottentots), and Bergdamara—were the original inhabitants of South Africa, primarily hunters and gatherers, but few now remain. In the seventeenth century, there were still perhaps 200,000 speakers of the Cape Khoikhoi dialects between the Cape and the Transkei, but disease and expropriation of grazing land by European settlers destroyed the Cape Khoikhoi's economic base and language. Now, a large number of Khoisan languages with usually a few hundred speakers each are scattered across the Kalahari region of southern Africa, including northern South Africa. At present, the total population may be estimated to be around 200,000 (Vossen 2000:131).

While the San (Bushmen) and the Khoi (Hottentots) lived in southern Africa possibly as far back as the Stone Age, the Nguni, whose descendants now comprise most of the country's black majority, arrived in South Africa by the third century AD. Up to

the fifteenth century, the Nguni occupied predominantly the northern, eastern, and south-eastern areas of South Africa, while the San and the Khoi occupied the southwestern regions from the Cape to the Orange River. By the nineteenth century, the Nguni and European settlers in South Africa pushed the Khoisan speakers into the Khalahari Desert. Members of the isinguni (the Nguni language group) include the Ndebele, Ngoni, Swazi, Xhosa, Zulu, and some smaller peoples, and speak closely related languages.

Currently, there are 31 main languages in South Africa, of which 27 are living. The extinct languages, all Khoisan languages, are Korana, Seroa, Xam, and Xegwi. The nearly extinct languages are Ng'huki and Xiri; both these are Khoisan languages. Fanagolo is a second language without mother-tongue speakers. The living languages include Afrikaans, Birwa, Chinese (Mandarin), English, Fly Taal, Gujarati, Hindi, Nama, Ndebele, Oorlans, Ronga, Northern Sotho, Southern Sotho (Sesotho), South African Sign Language, Swahili, Swati, Tamil, Tsonga, Tswa, Tswana, Urdu, Venda, Xhosa, and Zulu. Other minor languages include Dutch, German, Greek, Italian, Portuguese, French, and others. Few people in South Africa are monolingual, most are bilingual, and many are multilingual to varying degrees. Several languages are partially intelligible with other languages. For example, Tswa is partially intelligible with Ronga and Tsonga.

Population distribution is extremely uneven—more than two thirds live in the eastern third of the republic and in the southern Cape. Europeans have a widespread geographical distribution, but more than 80% reside in towns. Relatively few Africans are resident in the western Cape, and more than 60% continue to reside in those rural areas that comprised the formal tribal reserves. These extend in a great horseshoe along the southeastern coast and up to the Northern Province and then southwestward to the northeastern Cape. The 'Colored' population are mainly resident in the Cape, and the Asian population is concentrated largely in KwaZulu/Natal and the Witwatersrand.

African languages are unevenly distributed through the population (see, for example, ANC Constitutional Committee 1992). Ndebele is spoken mostly in northern and southern Transvaal. Setswana is the language of the Tswana people, spoken in parts of the Western Transvaal and Northern Cape, as well as a small area within the Orange Free State. Shangaan is spoken in Northern and Eastern Transvaal. Northern Sotho is the group of Sotho dialects spoken in the northern parts of South Africa. South(ern) Sotho is the group of Sotho dialects spoken in Lesotho and Orange Free State, as well as western Transvaal. Siswazi is the language of the Swazi people, spoken mostly in the Kingdom of Swaziland and in the Eastern Transvaal. Tshivenda or

Chivenda is the language of the Venda people, spoken mainly in the Soutpansberg region of the Northern Transvaal. Isixhosa is the language of the Xhosa people traditionally living in what is now the Eastern Cape Province. IsiZulu is the Zulu language; Zulus live mainly in KwaZulu-Natal, but isiZulu is spoken in the Orange Free State and the Eastern Transvaal, as well as Natal.

English was declared the sole official language of the Cape Colony in 1822 (replacing Dutch). Then, on the formation of the Union of South Africa in 1910, English was made the official language together with Dutch (which was replaced by Afrikaans in 1925). Afrikaans is a variant of Dutch spoken by the seventeenth-century colonists, with some lexical and syntactic borrowings from Malay, Bantu languages, Khoisan languages, Portuguese, and other European languages. After the establishment of the republic of South Africa in 1961, this policy of two official languages (English and Afrikaans) continued. The first democratic elections were held in 1994. The Interim Constitution of the new South Africa was adopted in 1994, establishing a federal, democratic, multiparty system, and identifying several aspects of education and language issues. In section 3, subsection 9c and 10d, the Interim Constitution made special mention of the 11 official languages. Eleven languages were accorded official status. The official languages in the 'new South Africa' are Afrikaans, English, isiNdebele (Ndebele), Sesotho sa Leboa, Sesotho (southern Sotho), seSwati, Xitsonga (Tsonga), Setswana (Tswana), Tshivenda (Venda), isiXhosa (Xhosa), and isiZulu (Zulu).

The new constitution of May 8, 1996 guarantees that the government will promote and protect African languages. It stated that the Pan South African Language Board be established to promote the 11 official languages as well as minority languages such as German, Portuguese, and the Indian languages. President Thabo Mbedi notes the progress in the establishment of the Commission for the Protection and Promotion of Language, Cultural, and Religious Rights as enshrined in South Africa's constitution (Mbeki 1999). Several researchers (e.g. Crawhall 1999) criticize the current government language policies for not sufficiently promoting and protecting the indigenous languages of the San, Nama, and the Khoe. The few surviving San and Khoi languages are all at risk of dying out in the next generation; land and natural resource access are essential for maintaining the languages of hunter–gatherers and traditional pastoralists (Crawhall 1999).

While linguistic pluralism is a national objective in democratic South Africa, many people engage in the monolingual practice of English only (Webb 1999). Prabhakaran (1998), for example, describes the complex process of an intergenerational language shift from Telugu, an Indian language, to English. He argues that the main causes of this shift are the dominant official status of the English language, the government language policy, the assimilation between the Andhras and the Tamils, and political factors such as the uprooting of the well-settled Andhra community by the Group Areas Act. English remains the main means of communication in South Africa's urban areas.

Estimates based on the 1991 census (Schuring 1993) indicate that approximately 45% of the South African population have a speaking knowledge of English. The majority of the population speak an African language as home language. Only about 10% of the population cite English as a home language. Of this figure, about one in three English speakers come from ethnic groups other than the white one (in proportionally descending order, from the South African Indian, Colored, and Black ethnic groups) (Lanham 1996). There is presently considerable (and overlapping) variation in the manifestation of English in South Africa. The variety of accents is a consequence of both the varied regional origins of groups of native English speakers who came to Africa at different times, and the variety of the mother tongues of the different ethnic groups. For example, one can distinguish between 'ethnic varieties' such as 'Colored', Black, South African Indian, Afrikaans English, and White South African English (Lanham 1996). And among white English speakers, there has been a traditional threefold distinction between 'conservative', 'respectable', and 'extreme' South African English (Lanham 1982).

Language planning, and particularly the role of the English language in South Africa's linguistic future, receives much debate. Discussions continue on the practicality of the language policy, the multiple vs. single language debate, 'tribalism', and the meaning of language and its role in identity (Finchilescu and Nyawose 1998). For example, Titlestad (1996) argues that the English language will inevitably play a powerful, leading role as an important *lingua franca* because of its role internationally. Since English is the international and intranational language of communication and many South Africans already speak English, he argues, market forces will play a role in ensuring the growth of English in South Africa. However, Webb (1996) maintains that in the interests of serving the needs of speakers of all languages in terms of the constitutional principle of multilingualism, language planning should be conducted from an Afrocentric view where all languages have equal rights.

References

Alexander, N. 1990. The language question. Critical choices: an agenda for South Africa in the 1990s, ed. by R. Schrire. Cape Town: Oxford University Press.

ANC Constitutional Committee. February 1992. Ten proposed regions for a united South Africa. Cape Town: CDS.

Crawhall, N. 1999. Going to a better life: perspectives on the future of language in education for San and Khoe South Africans. International Journal of Educational Development 19 (4–5). 323–35.

Finchilescu, G., and G. Nyawose. 1998. Talking about language: Zulu students' views on language in the New South Africa. South African Journal of Psychology 28(2). 53–61.

Greenberg, Joseph H. 1950. Studies in African linguistic classification. VI: the Click languages. Southwestern Journal of Anthropology 6(3). 223–37.

———. 1963. The languages of Africa. The Hague: Mouton.

Grobler, E., K.P. Prinsloo, and I.J. Van der Merwe (eds.) 1990. Language atlas of South Africa. Pretoria: HSRC.

Guldemann, Tom, and Rainer Vossen. 2000. Khoisan. African languages: an introduction, ed. by Bernd Heine and Derek Nurse. Cambridge: Cambridge University Press.

de Klerk, Vivian (ed.) 1996. Focus on South Africa. Amsterdam: John Benjamins Publishing Company.

Lanham, L.W. 1982. English in South Africa. English around the world: focus on Southern Africa, ed. by R. Bailey and M. Gorlach. Ann Arbor: University of Michigan Press.

Lanham, L.W. 1996. The pronunciation of English in South Africa. A dictionary of South African English on historical principles. Oxford: Oxford University Press in association with the Dictionary Unit for South African English.

Mbeki, Thabo. 1999. Protection and promotion of languages. Presidents and Prime Ministers 8(4). 26.

Titlestad, Peter. 1996. English, the constitution, and South Africa's language future. In de Klerk.

Vossen, Rainer. 2000. Khoisan languages with a grammatical sketch of Ani (Khoe). Areal and genetic factors in language classification and description: Africa south of the Sahara, ed. by Petr Zima. Muenchen: Lincom Europa.

Webb, Vic. 1996. English and language planning in South Africa: the flip-side. In de Klerk.

Webb, Vic. 1999. Multilingualism in democratic South Africa: the overestimation of language policy. International Journal of Educational Development 19(4–5). 351–66.

CAMILLA M. COCKERTON

South America: Argentina

Argentina, at the southern end of South America, has an area of 3,761,274 square kilometers and a population of 36,260,130 (2001 census). It is estimated that at least 35 different languages were spoken in this territory at the time of the first European contact in the early 1500s. In addition to Spanish, the official language, approximately 12 indigenous languages and a number of immigrant languages are also spoken at present.

Indigenous Languages

The indigenous languages of Argentina are classified into eight linguistic groups: Quechua, Tupi-Guarani, Guaycuruan, Matacoan, Lule-Vilela, Araucanian, Chon, and the unclassified language Yahgan. (The numbers of speakers of these languages are estimates, since no official record for the number of speakers for any of the languages exists.)

Quechua

Two varieties of Quechua are spoken in Argentina. One is the variety spoken mostly by migrants from Bolivia, with an estimated 500,000–850,000 speakers. The second is the Quechua spoken in Santiago del Estero, a province in the center of Argentina, known there as Quichua Santiagueño. Santiago del Estero Quechua was brought to this area by Quechua speakers who accompanied Spanish expeditions in the early years of conquest and who settled in the area. There are currently an estimated 60,000–100,000 speakers.

Tupi-Guarani

There are three Tupi-Guarani languages currently spoken in Argentina: Chiriguano, Tapiete, and Guaraní. Chiriguano is spoken by 15,000–21,000 people in the province of Salta, in northern Argentina. (There are approximately 50,000 speakers of Chiriguano in Bolivia.) This language is also spoken by a group called Chané, who at present speak a variety called Chiriguano-Chané. It is claimed that the Chané originally spoke Chané, a language believed by some to be of the Arawakan family. They were enslaved by the Chiriguano people, and eventually shifted to Chiriguano. The variety spoken by the Chané shows traces of the original Chané language, although their language has not been well studied yet. Tapiete (or Ñandeva) is spoken by approximately 400 people in the province of Salta. (There are approximately 3,200 speakers of Tapiete in Paraguay.) It is sometimes described as a variety of Chiriguano with a considerable amount of borrowings from other languages. Guarani is mostly spoken in Paraguay and Eastern Bolivia, but there are at least three varieties of Guarani spoken in Argentina mostly in the provinces of Formosa, Corrientes, and Misiones. Guaraní Correntino or Corrientes Guaraní is spoken mostly in the northeastern part of the country, with a considerable number of migrants in larger urban areas such as Rosario, Córdoba, and Buenos Aires. A second variety is known as Mbya, and is spoken in the province of Misiones mostly by people in rural areas. The third variety is

Paraguayan Guaraní, spoken by Paraguayan immigrants in northeastern Argentina, particularly in provinces closer to the Paraguayan border, and in large urban centers. The estimated number of speakers of the three varieties of Guarani ranges from 100,000 to 1,000,000. This is due partly to the lack of data on the Argentinian varieties, and in part due to the constant, undocumented influx of illegal immigrants from Paraguay.

Guaycuruan

There are four languages that belong to the Guaycuruan language family in Argentina: Mocoví, Pilagá, Toba, and the now extinct Abipón, in the Chaco region. (There is also one other Guaycuruan language, Kadiwéu, spoken in Matto Grosso do Sul, Brazil.) Mocoví is spoken by approximately 5,000 people in the provinces of Santa Fé and Chaco. Pilagá has an estimated 4,000 speakers in Formosa, mostly along the Pilcomayo River. Toba is the Guaycuruan language with the greatest numbers of speakers, 25,000, mostly in the provinces of Chaco and Formosa, but also as migrant groups in Santa Fé, Buenos Aires, and Salta. The Abipón were the first Guaycuruan Indians to settle in missions, known as reducciones, mostly in Santa Fe province; their language became extinct in the early twentieth century. Two other languages mentioned in several colonial sources are also believed to belong in the Guaycuruan family: Payaguá and Charrúa. However, the only extant records of these languages are brief word lists, and the evidence for their classification as Guaycuruan languages is very weak.

Matacoan

The Matacoan family (also known as *Mataguayo* or *Mataco-Mataguayo*) has three languages, spoken in northern Argentina: Wichí, Chorote, and Chulupí. (The fourth Matacoan language is Maká, spoken in Paraguay.) Wichí is commonly known as Mataco, but this is a pejorative term rejected by speakers of the language. It has between 40,000 and 60,000 speakers in Chaco, Salta, and Formosa provinces. Chorote is spoken in Salta by approximately 2,000 people (and an estimated 500 others in Paraguay). Chulupí (also known as Nivaclé and Ahluhlay or Ashlushlay) is spoken by anywhere from 200 to 1,200 people in Salta — estimates vary widely (and perhaps as many as 18,000 in Paraguay). In Argentina, they mostly live in communities mixed with Chorote and Wichí speakers.

Lule-Vilela

This is a small proposed language family, with only two languages: Lule and Vilela. Lule, believed to be a dialect of the Tonocoté language referred to in early documents, was spoken mostly in Tucumán, western Salta, and northwestern Santiago del Estero. It became extinct in the nineteenth century. It is believed to be related to Vilela, a language spoken in the provinces of

Salta and Chaco by a small number of speakers until the mid-1960s, now believed to be extinct.

Araucanian

One language belongs to the Araucanian family, Mapudungun (also known as Mapudungu, Mapuzungun, Araucano, or Mapuche, although the last term refers to the speakers, from *mapu* 'earth' and *che* 'people'). It is spoken by approximately 40,000 people in Argentina in the provinces of Neuquén, Río Negro, La Pampa and Chubut, and various communities in Buenos Aires province. It is also spoken in Chile. Mapudungun spread to Argentina from Chile through the southern Andes before European arrival. There was sporadic contact between the earlier groups in Patagonia and the Mapuche, mostly for commercial purposes, but in the early seventeenth century, large numbers of Mapudungun speakers crossed the Andes spreading into Patagonia and the central plains, known as the Pampas, forcefully occupying Tehuelche territory, imposing their language and culture. This 'Araucanization of the Pampas' lasted until the late nineteenth century, with the violent imposition by the Argentinian government of the authority on the indigenous groups of the area.

Chon

The languages of the Chon family were spoken in Patagonia in southern Argentina and the southern islands. They can be divided into two groups: Island Chon and Tehuelche, reflecting more a geographic grouping than a linguistic one. Two languages belong in the Island Chon group: Haush and Selk'nam, which were spoken in the island of Tierra del Fuego. Haush, or Manek'enk, was spoken in the southeastern tip of the island, and became extinct in the early twentieth century. Only brief vocabularies of Haush remain, dating from the late nineteenth century. Selk'nam, or Ona was spoken in the central and northwestern parts of the island. Although the language was not extensively studied, it is better documented than Haush. By 1973 there were only three speakers of Selk'nam, and the language is now extinct. Some believe Haush to be but a variety of Selk´nam, and not a distinct language. The second group, Tehuelche, was spoken by groups in Patagonia and the Pampas. Four varieties of Tehuelche are reported: Northern Tehuelche, also known as Gününa-küne, was spoken from the Colorado River to the Chubut River. It became extinct in the 1960s. A second variety of Tehuelche is believed to have been spoken in the southern Pampas, from Eastern Neuquen to the Colorado River, but very little documentation exists for the language (mainly toponyms and a few phrases from travelers). Southern Tehuelche was spoken from the Chubut River to the Magellan Strait. Two varieties are reported: Teushen,

now extinct and of which very little is known (briefly reported in vocabularies dating from 1780 to 1900), was spoken between the Chubut and Santa Cruz Rivers. The second variety, known as Tehuelche or Aonek', was spoken from the Santa Cruz River to the Straits of Magellan. Only two speakers of Tehuelche are alive today, but they do not use the language regularly and only remember it partially.

Yahgan

Yahgan, also known as Yámana, was spoken in the southern coast of Tierra del Fuego and the southern islands of the area. The language is not part of the Chon family, and its genetic affiliation is still unclear. The language became extinct in Argentina in the early twentieth century. Today, there is now only one speaker of Yahgan left in Chile, an elderly woman. The only documentation of the language was provided by the first missionary to settle in the area, the Englishman Thomas Bridges.

Other languages

Argentina was first settled by Spanish colonizers in the early 1500s, and while the population grew slowly in the first two centuries, the late nineteenth and twentieth centuries witnessed a surge in immigration, mostly from Europe. Although the Argentinean government provided incentives for the immigrants to settle in the interior of the country, the majority settled in Buenos Aires and surrounding areas. Most first-language speakers of these immigrant languages are first-generation immigrants, and by the second or third generation speakers shift to Spanish, learning the immigrant language (if at all) only as a second language. Two particularly interesting cases are those of Welsh and Korean. A large group of Welsh immigrants settled in the Patagonia, on the Chubut River in 1865 as a result of an agreement between the Argentinean government and English representatives. They were able to maintain their language well into the twentieth century,

gradually shifting to Spanish, after decades of pressure from the Argentinean authorities, so that by the 1940s Welsh was no longer spoken among members of the Welsh community. Korean speakers represent the latest 'wave' of immigration, mostly as a result of a diplomatic agreement between the Argentinean and Korean governments. The largest number of Korean speakers came to Argentina between 1965 and 1989, with most between 1984 and 1989. Patterns of language maintenance among Korean speakers reflect the same trends as other immigrant languages, with second- and third-generation speakers already shifting to Spanish. The immigrant languages with a considerable number of speakers include Italian, German, English, Welsh, Ukrainian, Polish, Lithuanian, Vlach Romani, Arabic, Japanese, and Korean.

References

Bridges, Esteban Lucas. 1948. Uttermost part of the earth. London: Hodder & Stoughton.

Censabella, Marisa. 1999. Las lenguas indígenas de la Argentina, Una mirada actual. Buenos Aires: EUDEBA.

Cooper, John M. 1946. The Patagonian and Pampean tribes. In Steward.

Kaufman, Terrence. 1994. The native languages of South America. Atlas of the world's languages, ed. by Christopher Moseley and R.E. Asher. London, New York: Routledge.

Mera, Carolina. 1998. La inmigración coreana en Buenos Aires: multiculturalismo en el espacio urbano, Buenos Aires: EUDEBA.

Métraux, Alfred. 1946. Ethnography of the Gran Chaco. In Steward.

Serrano, Antonio. 2000. Los Aborígenes Argentinos, Síntesis Etnográfica. Córdoba: Ediciones Librería Paideia.

Steward, Julian H. (ed.) 1946. Handbook of South American Indians, Vol. 1. Washington: Smithsonian Institution, Bureau of American Ethnology, Bulletin 143.

Williams, Glyn. 1991. The Welsh in Patagonia, the state and the ethnic community. Cardiff: University of Wales Press.

Veronica M. Grondona

See Also **Quechua and Andean Equatorial Languages; Wayampi and Tupí-Guaraní Languages; Spanish**

South America: Brazil

Occupying nearly half the continent of South America, Brazil is one of the largest and most populous countries in the world, and its inclusion of a major portion of the Amazon Basin adds to its enigmatic and diverse character. Brazil is home to a large number of little-known indigenous languages from a variety of language

families, as well as to several genetic isolates—languages that do not seem to be related to any other known languages. The status of Brazil is also unique in South America as the only country recognizing Portuguese rather than Spanish as its official language. Finally, the importation of African slaves into the region, as well as continued immigration by a variety of ethnic groups, has created an interesting and complex linguistic situation in Brazil.

Indigenous Languages

With no written records documenting precolonial Brazil, estimates of the indigenous population at the time of the Portuguese arrival in 1500 have ranged between one and six million. As a consequence of war and the introduction of such European diseases as smallpox and influenza, the indigenous population was drastically reduced. Today, there are about 200,000 Indians in Brazil.

Despite the loss of languages that certainly occurred due to the loss of human life, about 170 different Amerindian languages remain in the region today. Because little is known about many of these languages, and because contact among the various languages has been great, a great number of classifications have been proposed for the languages of Brazil. One of the most commonly accepted proposes five major families—Arawak, Carib, Tupi, Macro-Je, and Pano—as well as several smaller families and some genetic isolates. The highest density of indigenous languages and indigenous speakers in South America is in the Amazon basin in the northern region of Brazil. Tupi and Arawak are two of the most important of these languages today, and Tupi has contributed greatly to the word stock of Brazilian Portuguese. Sharing a number of similarities with Tupi, Carib is spoken primarily in the northern region of Brazil as well as in the northern South American countries of Colombia, Venezuela, Guyana, Belize, and Guatemala. The two language families are not only thought to have so many similarities due to genetic affiliation, but also because of a great deal of contact among some of the languages of the two families. Although some of the indigenous languages of South America build sentences from very simple words, most tend to favor highly complex word structures and most tend to prefer suffixation over prefixation. In addition to interest in how language families relate to one another genetically, there is also great interest in how languages within families are related to one another geographically, especially those associated with the Amazon Basin, as many families are distributed disparately throughout the area. Additionally, due to a history of slavery among the tribes of the region, multilingualism has been very common among the Indian tribes of the Amazon basin.

Despite some effort to maintain indigenous languages, many languages have been lost due to assimilation into mainstream society. Languages continue to be lost in Brazil as indigenous peoples are assimilated into the mainstream culture and younger generations become monolingual speakers of Portuguese. While the linguistic diversity of Brazil, as with South America as a whole, is widely recognized, description of these languages has been insufficient for a number of reasons such as the bureaucratic difficulties of working with indigenous peoples, inadequate funding, and limited linguistic training.

Portuguese

The Portuguese were the first Europeans in Brazil, arriving in 1500, several years after Christopher Columbus' first voyage to the Caribbean. Although the Spanish, led by Columbus, were the first Europeans to arrive in the New World, the Portuguese had always dominated the seas in terms of long-distance travel. By the end of the fifteenth century, the Portuguese had already succeeded in sailing around Africa to reach India, establishing an important trade route. The discovery of the Americas by the Spanish no doubt threatened Portugal's dominance of the seas, and the Portuguese soon sought to establish their own colonies in the New World.

The status of Brazil as the only Portuguese-speaking country in South America has to do with the ways in which the Portuguese and the Spanish made use of their colonized land. A treaty between Portugal and Spain in 1494 divided South America into two parts, giving Spain control of the western side and Portugal control of the eastern side. In Brazil, the main source of wealth became the production of sugar cane, while in the Spanish colonies, the main source of wealth was the extraction of precious metals, most notably gold and silver. Sugar cane production required a highly organized society in Brazil, with large pieces of land and numerous slave workers imported from Africa. The search for gold and silver, on the other hand, could be done by smaller groups of people, and its success often relied on chance. Thus, while the Portuguese colonization of Brazil was fairly organized, the Spanish colonization of the rest of South America was more fragmented and lacked a single centralized government.

There was never any question that Portuguese would be the official language of Brazil; what was under debate was which Portuguese would be used. The popular Portuguese of Brazil has always been markedly different from European Portuguese. The Brazilian people do not consider themselves direct descendants

of the Portuguese colonists but rather as a mixture of Portuguese, Indian, and African people, and the language of Brazil has always reflected this mixture of cultures. By the nineteenth century, people began to debate the existence of a 'Brazilian language', and writers such as José de Alencar reacted by using words and expressions they felt reflected Brazilian rather than Portuguese culture. This trend continued into the twentieth century, especially in the modernist movement of the 1920s. Nevertheless, Standard Brazilian Portuguese tends to follow norms inherited from European Portuguese, even while allowing for some differences in vocabulary and spelling. For instance, in European Portuguese, the subject (pronoun) of the sentence is frequently dropped, and phrases generally avoid beginning with an object pronoun. Rather, the object pronoun is placed after the verb, as in *esperou-me* 'he or she waited for me' (lit. 'awaited-me'). In colloquial Brazilian Portuguese, however, it is more common to use subject pronouns, and object pronouns are almost always placed before the verb (or not used at all in the case of the third person); thus, *ele/ela me esperou* 'he or she waited for me'. (lit. 'he or she me awaited') However, forms such as *esperou-me* are frequently encountered in texts, and such forms are the ones taught in schools. Although phrases beginning with an object pronoun, such as *me esperou*, are not uncommon in speech, grammarians consider them incorrect usage.

Differences in vocabulary and pronunciation between European and Brazilian Portuguese tend to be less stigmatized than grammatical differences. Many lexical items are different in the two varieties, such as Brazilian *meia* and European *peúga* 'sock'; others have different spellings and/or pronunciations, such as Brazilian *fato* and European *facto* 'fact'. It is pronunciation that most clearly marks speech as either Brazilian or European, regardless of the level of formality. Some highly noticeable differences include the pronunciation of final -e in words like *nove* 'nine' as [i] in Brazilian Portuguese and as either schwa or silent in European Portuguese and the palatalization of /t/ and /d/ before /i/ in most regions of Brazil (i.e. *ti* is pronounced as [tʃi], *di* as [dʒi]).

There are also many varieties of Portuguese in Brazil with every major city noted for having its own accent. Rio de Janeiro, for example, is characterized by its use of alveopalatal s and z in the syllable-final position. A word such as *nós* 'we', then, is normally [nɔʃ] in Rio and surrounding areas, while in most other cities in Brazil, it is typically [nɔs].

African Languages

The ancestors of African Brazilians arrived as slaves beginning in the middle of the sixteenth century and ending with the end of the slave trade in 1850. More than two million slaves were imported into Brazil to provide labor on the sugar plantations and in the gold mines, with the greatest numbers being brought in during the first part of the nineteenth century. This population greatly increased that of Brazil during this period. The varieties of African languages that existed in Brazil were influential in the formation of various Brazilian creoles and had some effect on Brazilian Portuguese; however, these languages have generally been lost in Brazil over the years. Yoruba and other African languages, however, continue to be used in the religious ceremonies of the Candomblé societies.

Creole Languages

The great variety of indigenous languages in Brazil and language contact between them and the early Portuguese, as well as the importation of African slaves, resulted in the development of a variety of pidgins and creoles in Brazil. Lingua Geral, a trade language created by Tupi Indian nations along the coast, for instance, was popular until the end of the seventeenth century and was still used in some isolated areas, particularly in the Amazon Basin, until the late nineteenth century. Brazilian Creole Portuguese is still spoken to some extent, mostly by rural Brazilians of African descent in the north. Although these various contact languages were tolerated and even viewed as useful during the early colonization of Brazil, the rise of Portuguese as the official language of the country, and as the lingua franca in contact situations, has done a great deal to reduce the creole-speaking population.

Other European Languages

Immigration into Brazil by non-Portuguese Europeans began around 1850, primarily involving Italians and Germans settling in the subtropical regions near the Atlantic coast, especially in the states of São Paulo, Paraná, Santa Catarina, and Rio Grande do Sul; immigration by these groups generally ended by the time of World War II. These languages are still used in some of the more homogeneous of these communities to some extent, particularly in religious ceremonies. The Japanese began arriving and settling in approximately the same region around the turn of the twentieth century, and their language is still used to a great extent in Brazil. English and French, and to a lesser extent German and Italian, are taught as second languages in schools.

Illiteracy

In addition to the loss of indigenous languages in Brazil, another significant problem of note is the high

illiteracy rate, estimated in 2002 as 16.7% of the population. Although there are numerous literacy programs available in Brazil, especially in churches, and literacy has become much more widespread since the early 1900s, illiteracy is in reality only a small part of a larger problem. People in the lowest classes are not always encouraged or even expected to go to school, and the drug trade and other crimes often provide a much faster way to make money than getting an education.

References

Aikhenvald, Alexandra Y., and R.M.W. Dixon. 1999. Other small families and isolates. In Dixon and Aikhenvald.

Bortini-Ricardo, Stella Maris. 1985. The urbanization of rural dialect speakers: a sociolinguistic study in Brazil. Cambridge: Cambridge University Press.

Dixon, R.M.W., and Alexandra Y. Aikhenvald (eds.) 1999. Introduction. The Amazonian languages. Cambridge: Cambridge University Press.

Greenberg, Joseph H. 1987. Language in the Americas. Stanford: Stanford University Press.

Key, Mary Ritchie. 1979. The grouping of South American Indian languages. Tubingen: Narr.

Melo, Gladstone Chaves de. 1971. A Lingua do Brasil. Rio de Janeiro: Fundacao Getulio Vargas.

Payne, Doris L. (ed.) 1990. Amazonian linguistics: studies in lowland South American languages. Austin: University of Texas Press.

Rodrigues, Aryon D. 1985. The present state of the study of Brazilian Indian languages. South American Indian languages: retrospect and prospect, ed. by Harriet E. Manelis Klein and Louisa R. Stark. Austin: University of Texas Press.

Thomas, Earl W. 1974. A grammar of spoken Brazilian Portuguese. Nashville: Vanderbilt University Press.

LAMONT ANTIEAU AND MICHAEL COLLEY

See also **Arawak; Carib and Cariban Languages; Pidgins and Creoles; Wayampi and Tupí-Guaraní Languages**

Southeast Asia

A region comprising both the Indochinese and Malay peninsulas, as well as several island groups in the area, Southeast Asia is home to a great number of people and cultures. Bordered on the north by the Yangtze River in China, on the east by the South Pacific Ocean, on the south by the Indian Ocean, and on the west by the Indian Ocean, the Bay of Bengal, and the subcontinent of India, Southeast Asia includes the countries of Brunei, Myanmar (formerly Burma), Cambodia (Kampuchea), Indonesia, Laos, Malaysia, Singapore, Thailand, Vietnam, and the Philippine Islands, as well as the southernmost region of China. Owing in part to the diverse geographical makeup of the area and the lengthy and intricate historical relationships of its inhabitants, Southeast Asia is a region of great linguistic variety and complexity.

Although more than a thousand languages are spoken in Southeast Asia, five language families are particularly well represented in the region: Sino-Tibetan, Mon-Khmer, Tai, Miao-Yao (also known as the Hmong-Mien), and Austronesian. While such major languages of the Sino-Tibetan family as Chinese and Tibetan are spoken primarily in China and the Himalayas, speakers of these languages can also be found in great numbers throughout Southeast Asia, particularly in those countries sharing a border with China. In addition to being widely spoken in Southeast Asia, Chinese has also had a great influence on many

languages in the region, particularly with respect to vocabulary. Due to China's long domination of Vietnam from 111 BC to AD 939, Chinese has probably had the greatest influence on Vietnamese, most notably in its legal and medical terminology. A number of other Sino-Tibetan languages are spoken almost exclusively in countries of Southeast Asia, including Burmese, the official language of Myanmar, a country of over 35 million people.

The Mon-Khmer language family is another important family in Southeast Asia and includes Khmer, which is the official language of Cambodia and has approximately seven million speakers, and Mon, which was an important language in earlier times, but has been reduced to about one million speakers. Just as Chinese has had a great influence on the languages of Southeast Asia, so too has Khmer, particularly during the Angkor period, which lasted from approximately the ninth to the fifteenth centuries AD. As a result of Khmer dominance during this period, Khmer vocabulary is well attested in the lexicons of many languages of Southeast Asia, including Thai and Lao. Vietnamese, which is the official language of Vietnam and is spoken by about 80 million, is also classified by most scholars as a member of the Mon-Khmer family; however, this classification has been somewhat controversial, mostly due to the phonemic status of tone in Vietnamese, a trait not shared by other Mon-Khmer languages. This

difference led earlier scholars to classify Vietnamese as a Tai language; however, Haudricourt (1954) accounts for the use of tones in Vietnamese as the result of an evolutionary process in which Vietnamese acquired tones from other languages in Southeast Asia, notably Chinese, via language contact. Linguists often classify the Mon-Khmer family as a member of the Austro-Asiatic family of languages together with the Munda family of India.

Another language family of great importance in Southeast Asia is the Tai family, a group of tonal languages spoken primarily in Thailand, Laos, and southern China. The two major languages of the Tai family are Thai, which is spoken by about 50 million people and is the official language of Thailand, and Lao, which is spoken by about 10 million and is the official language of Laos. Other languages in the Tai family include Shan, a language spoken by over two million in Myanmar, and Chuang, which, with over 15 million speakers, is the language of the largest ethnic minority in China.

Languages in the Miao-Yao family are spoken in isolated areas of southern China and in the northern regions of Laos, Thailand, and Vietnam. The main languages of the group are Miao, which is spoken by about two and a half million people, and Yao, spoken by about one million. Many scholars consider the Miao-Yao family a branch of the Sino-Tibetan family, although it has also been proposed that the family is a branch of either the Tai or the Mon-Khmer family.

The Austronesian languages, formerly called Malayo-Polynesian, are spoken on the Malay Peninsula and on most of the islands to the southeast of mainland Asia. Distribution of the family's languages actually extends far beyond the region of Southeast Asia with member languages spoken from Madagascar to Easter Island and from Taiwan to Hawaii, as well as in and around New Zealand. The family comprises approximately 200 million speakers of at least 500 languages. In Southeast Asia, the principal Austronesian languages are Indonesian, which is the official language of Indonesia and is spoken by about 20 million, and Malay, spoken by about 10 million and the official language of Malay. Some linguists have claimed that Indonesian and Malay are essentially the same language, differing only in orthography and political identity. Another important language of the Austronesian group is Tagalog, which has been the national language of the Philippine Islands since 1962 and has about 15 million speakers.

Despite the great number of differences between the various languages spoken in Southeast Asia, some scholars have pointed out a number of commonalities, suggesting that the languages share a genetic relationship as well as an areal one (see e.g. Matisoff 1978). These shared characteristics include similar patterns of nominalization, compounding, and the use of classifiers rather than morphemes to mark gender. Additional support for this position is that discourse patterns tend to be similar in various Southeast Asian languages. However, the often intense and diverse contact in the region makes it as difficult for linguists to determine whether characteristics have been borrowed by languages as it is to identify with precision the classification of the various languages. In addition to the problem created by contact among neighboring languages for hundreds of years, the great influence over the entire region by such historical powers as Chinese and Khmer, as well as the influence of languages and cultures from outside Southeast Asia, tends to blur the distinction between languages. Numerous words, for instance, have been borrowed into many Southeast Asian languages from Sanskrit, while others have been borrowed from Pali due to the adoption of Buddhism by some Southeast Asian cultures.

Languages well outside Southeast Asia, especially a number of European languages, have also had an impact on the languages of the region. French has been an extremely influential language, particularly in Vietnam and Cambodia, countries that were under the rule of the French from the mid-nineteenth to the mid-twentieth century. During this time, French was recognized as the official language of both countries, serving primarily as the language of government and education. Although both Vietnam and Cambodia became independent of French rule shortly after World War II, the bonds between France and Southeast Asia have not been completely severed as evidenced in the continuing migration of natives of Southeast Asia to France. The influence of the French on various Southeast Asian languages is most readily apparent in their lexicons, especially with respect to words relating to the government and the military. Spanish is also of some significance in Southeast Asia due to its long domination of the Philippines from the mid-sixteenth century until the end of the nineteenth century; the best evidence of Spanish is found in the lexicon of Tagalog, which includes hundreds of Spanish words. English is yet another European language that has had some influence on the region, primarily due to the American presence in the Philippines during the first half of the twentieth century, American involvement in the Vietnam War, and the role of English in international trade and diplomacy.

Contact between Southeast Asians and others has resulted in the formation of a great number of pidgin and creole languages in the area. As in other parts of the world, Portuguese played a crucial and early role in the creation of pidgins and creoles in Southeast Asia, particularly in Indonesia, which hosted a number of Portuguese-based creoles that are now all considered extinct. The Portuguese presence in the region, which

dates back to the fifteenth century, also led to the development of a number of other Portuguese-based creoles, including Burma-Siam Creole Portuguese, of which a small number of speakers can be found in Penang, and Papia Kristang, which is spoken in Malaysia and Singapore. Perhaps an even greater number of pidgins and creoles in Southeast Asia developed due to contact between speakers of Spanish and speakers of various languages on the Philippine Islands, where the number of Creole Spanish speakers stands at around 280,000. These contact languages include Bamboo Spanish and Kitchen Spanish, which are used primarily by Chinese shopkeepers in the Philippine Islands. Another contact language that developed in Southeast Asia, this time as a result of contact between speakers of French and Vietnamese, was Tay Boy, a Vietnamese-French pidgin spoken during the French domination of Vietnam from approximately 1860 to 1960. Contact between speakers of English and Vietnamese during the Vietnam War resulted in Vietnam Pidgin, which is now largely extinct; in Singapore, contact between English speakers and native inhabitants resulted in the formation of Singlish. There is also a mixed language in Southeast Asia called Davaoeno, which is a fusion of Tagalog and Visayan and is spoken by approximately 125,000 speakers in Davao.

One linguistic issue of great importance in Southeast Asia, as it is in Asia in general, concerns the various writing systems that have been implemented to represent the languages of the region. Many Southeast Asian languages, including Burmese, Cambodian, Lao, and Thai, are written in a Brahmi script that was imported from South India in the fourth or fifth century AD and has undergone some modification. Several languages, including Vietnamese and Indonesian, are now written in the Roman alphabet, which was introduced by missionaries in the region in the seventeenth and eighteenth centuries, and, in the case of Vietnamese, replaced an earlier system based on Chinese orthography. Still other languages, such as Malay, have incorporated some Arabic characters into their writing systems. Each of these writing systems presents a number of problems. Although the

Southeast Asian languages that use the Brahmi script have modified it to some degree to reflect the sounds of their own languages, the Indian language that it was borrowed from differed significantly from the Asian languages it is now used to represent. The result has been the extensive use of subscript and superscript symbols to represent vowels. Although some have pointed to it as an example of successful writing reform, Vietnamese shares in this problem and makes such wide use of diacritics that some scholars have remarked that it hardly constitutes true reform (see Hannas 1997). The problem that these systems share is the difficulty they create for printing processes, both because of the amount of paper required for printing the great number of diacritics as well as the extra time required to keyboard the extra symbols. These problems have typically been at the core of the debate over writing reform for Southeast Asian languages in modern times. Another has been that there are many minority languages in Southeast Asia, often spoken in the mountains of the region as opposed to the plains and valleys, for which there are no writing systems at all.

References

Dinh-Hoa, Nguyen. 1980. Language in Vietnamese society: some articles by Nguyen Dinh-Hoa. Carbondale, IL: Asia Books.
———. 1997. Vietnamese. Amsterdam: John Benjamins.
Hannas, Wm. C. 1997. Asia's orthographic dilemma. Honolulu: University of Hawaii Press.
Haudricourt, André-Georges. 1954. De l'origine des tons in viêtnamien. Journal Asiatique 242. 68–82.
Huffman, Franklin E. 1970. Cambodian system of writing and beginning reader with drills and glossary. New Haven: Yale University Press.
Jacob, Judith M. 1968. Introduction to Cambodian. Oxford: Oxford University Press.
Matisoff, James A. 1978. Variational semantics in Tibeto-Burman. Philadelphia: Institute for the Study of Human Issues.
Parkin, Robert. 1991. A Guide to Austroasiatic speakers and their languages. Honolulu: University of Hawaii Press.
Smith, Norval. 1995. An annotated list of Creoles, Pidgins, and mixed languages. Pidgins and Creoles: an introduction, ed. by Jacques Arends, Pieter Muysken, and Norval Smith, 331–74. Amsterdam: John Benjamins.

LAMONT ANTIEAU

Soviet Union

From its inception in 1917 to its ultimate breakup in 1991, the Soviet Union was a multilingual and multiethnic state. At its point of greatest expansion, it encompassed some 8,649,490 square miles with a total population of just under 286,000,000. The population is unevenly distributed among the ethnic groups and

languages, with some very large and some quite small in number. Russians comprise the largest single ethnic group, constituting just over 50% of the total population (or slightly more than 145,000,000) in the 1989 census. Russian was one of several Slavic languages spoken in the USSR, although by far the most widely used and it alone had the status of a national lingua franca. The Russian language has by far the largest number of speakers, with 81% of the population seeing itself as fluent in Russian, either as a first or second language.

Language and Ethnicity

The 1989 Soviet census cites approximately 130 ethnic groups (including indigenous and immigrant populations) speaking a total of 150 languages. (Linguists, however, estimate that there were actually closer to 200 languages spoken in the former USSR.) There is not, however, any one-to-one correspondence between language and ethnicity, as these figures show. In the USSR, all citizens aged 16 years and higher were required by law to declare their 'ethnicity' (*natsional'nost'*), which was officially recorded in each individual's (internal) passport. The census regularly asked questions about language use and ethnicity as well, but because language/dialect boundaries were often determined by social and political factors rather than linguistic ones, and the number of official languages and ethnic groups changed in accordance with changing political goals. For example, three Tungusic languages—Even, Evenki, and Negidal—were long considered to be dialects of a single language, although all are linguistically distinct, so that when they each acquired official recognition, there was a net increase in the number of ethnic groups. Accordingly, official Soviet statistics should be taken as providing only partial information about the linguistic map. Speaker population size and density vary significantly: some languages have millions of speakers, while others have only thousands or even hundreds of speakers. Some are in dense speaker communities in urban areas, while others are traditionally nomadic and have dispersed populations.

Linguistic Diversity

The many languages of the USSR can be classified into five large language families: Indo-European, Caucasian, Altaic, Uralic, and Paleo-Siberian. In addition, there are a number of language isolates (e.g. Ket, Nivkh, Yukagir). A number of Indo-European languages are native to territories of the USSR: the East Slavic languages (Russian, Belorussian, and Ukrainian); the Baltic languages (Latvian and Lithuanian); Armenian; a Romance language (Moldovan, which is minimally dis-

tinct from Romanian); Iranian (Tadzhik) as well as Germanic (German and Yiddish), and a few others. Of these genetic groups, Indo-European is the most well defined. The genetic relations of the Uralic languages (e.g. Finno-Ugric languages such as Estonian, Permic, Samoyedic, and Volgaic) are also clear. The Caucasian languages are divided into three or four subgroups, whose relations to one another are disputed. The Altaic languages are commonly divided into three language families (Mongolian, Tungusic, and Turkic), and although these three families constitute clear genetic groups, their relations to one another are the matter of much debate. Paleo-Siberian is an even more loosely defined category, as it groups together the languages of Siberia that do not belong to any of the other genetic groups (e.g. Altaic, Uralic, etc.). Two relatively small language families (Chukotko-Kamchatkan and Eskimo-Aleut) are usually placed in the Paleo-Siberian group.

The Soviet Union was organized into 15 union republics (or SSR, for Soviet Socialist Republic), which can be grouped into categories according to geography: the Baltics (Estonian SSR, Latvian SSR, and Lithuanian SSR); the Caucasus (Armenian SSR, Azerbajdan SSR, and Georgian SSR); Central Asia (Kazakh SSR, Kirgiz SSR, Tadzhik SSR, Turkmen SSR, Uzbek SSR); Slavic territory and Moldova (Belorussian SSR, Moldavian SSR, Russian SFSR, Ukrainian SSR); and the Russian Far East (RSFSR, primarily Siberia). Each Republic bears the name of its majority ethnic group, but it should be stressed that these are political, not ethnolinguistic divisions. Ethnic Russians lived in all republics in varying numbers. The Russian Republic was the largest in terms of territory and also had the highest percentage of ethnic Russians. Moreover, varying numbers of other ethnic groups lived in each of the republics, and some were dispersed over several. The Tatars, for example, with a total population count of 5.5 million in 1989, live scattered in 80 different regions of the former Soviet Union, with only 26.6% of their total numbers living in Tatarstan (Russian SFSR), another 17% in other regions of Russia, 7% in Uzbekistan, and so on.

Soviet Language Policy

The newly formed Soviet state faced illiteracy rates of nearly 100% in some areas, with an overall average literacy rate of only 28.4% for the entire country. In order to achieve their goal of rapid industrialization, the Bolshevik leaders required a fairly educated work force, which made a rapid increase in literacy one of their priorities. The sincerity of early Soviet language policy has been the subject of much debate, but Lenin's basic principle was unambiguous: no one language should be given the status of a state language

but rather, national equality and self-determinism must be promoted. The Declaration of Rights of the People of Russia (November 2, 1917) proclaimed equality for all people, with 'the free development' of the national minorities and ethnic groups inhabiting the USSR. All Soviet citizens were guaranteed education in their native language.

The literacy campaign was a fundamental part of the larger nativization or, literally, 'rooting' (*korenizacija*) policy, a policy intended to educate the indigenous peoples and move them into the workforce, especially into the Soviet administrative workforce. A major obstacle to achieving this goal was the overall low educational levels nationwide. Thus, the success of the nativization policy depended upon the ability of the government to educate its people. At the time of the formation of the Soviet Union, the majority of languages needed linguistic description and codification, and creation of a written form. Some linguistic regions (such as Georgia and Armenia in the Caucasus, and the Turkic-speaking Central Asia) had long-standing literary traditions, as did the Georgian and Armenian languages in the Caucasus and the Turkic languages in Central Asia. At the same time, many of the languages of the newly formed empire lacked written forms. It is estimated that at the time of the Bolshevik Revolution, only 13 languages on Russian territory had a literary norm, and only 19 had any kind of written form at all.

The Roman alphabet was initially chosen as the basis for all new writing systems and was also used for those languages not using Cyrillic, such as those using an Arabic or Mongolic script, although Georgian and Armenian were allowed to maintain their own alphabets. The use of Romanized script was short-lived, being phased out in favor of Cyrillic in the mid- to late 1930s and completed in the early 1940s. This transition to Cyrillic-based alphabets was part of a shift in language policy, which began after Lenin's death in 1924. A key date is 1934, when Stalin effectively ended the nativization campaign in an address to the XVII Party Congress. In March 1938, an official decree made the study of Russian compulsory.

Alphabet creation was just one step in the literacy campaign. Equally important was the creation of a standardized literary form for each of the targeted languages. Originally, emphasis was placed on phonetic spelling, which created problems due to vast dialectical differences for some languages, and due to the large influx of Russian borrowings as a result of the sociopolitical and economic changes. The creation of a new lexical inventory was an equally vital component of this larger campaign. Initially, it was argued that each of the national languages should have a complete inventory of all technical terms, created using language-internal resources wherever possible.

Ultimately, this did not occur, as the spheres of usage of many of the national languages were quite limited, making this native technical terminology superfluous. Where loans were concerned, the early policy was to maintain the pronunciation of the lending language. This policy was changed by the 1950s, when all loanwords were written in the original Russian form.

Despite the many controversial aspects of the Soviet literacy campaign, it did achieve rapid results. The literacy rate climbed dramatically from 44.1% in 1920 to 87.0% in 1939 to an official rate of 99.7% in 1979.

The mid-1950s witnessed yet another major change in language policy. Prior to this period, the national languages were the focus of Soviet language planning. From this time onward, the goal of Soviet language policy was to establish Russian as the language of the Soviet Union. The Khrushchev era (1953–1964) introduced the vision of a new Soviet people, united not only politically but also through the use of one language. Khrushchev declared Russian to be 'the second national language'. Although bilingualism was openly promoted, the very need for national languages (any language other than Russian) was questioned. The Khrushchev introduced the notion of the 'relative' importance of languages, and it became officially acceptable to view some languages as less viable than others: some languages were deemed unsuitable for development. Yet, Lenin's policies of language equality were not officially repudiated, and the Communist Party could invoke them to justify its own policies at any time.

This occurred along with a major change in education policy. Clause 19 of the Education Reforms of 1958–1959 stated that education in the mother tongue was no longer compulsory. By this time, instruction in the native language was offered for most languages with a written form at an elementary level, and at a secondary level for some. One impact of the Education Reforms was pressure to begin instruction in Russian from the earliest grades, and the native language was replaced by Russian in many schools. Even where native-tongue education continued, Russian was a compulsory subject.

The results of the Education Reforms varied throughout the country. Despite the shift in emphasis on Russian, languages spoken by larger populations may have actually gained some ground, in part due to the lessening of cultural restrictions under Khrushchev. This was the case in Central Asia, for example. In contrast, minority languages became seriously threatened as they were no longer used in schools, and publications in these languages were seriously cut back. Regardless of the local-level particulars, the major change of this period is that Russian became the official language of the USSR and occupied a central position in education and government.

This process continued under Brezhnev (1964–1982), in a greater move toward total Russification, with increasing pressure to make Russian the 'second mother tongue'. Official statistics show Russians to be largely monolingual (97%), while over 40% of the non-Russian population claimed itself to be bilingual, and by 1979 a total of 82% of the population claimed some knowledge of Russian. Translation work was primarily unidirectional, from Russian into the native language. The Brezhnev period is characterized by a steady increase in both the sheer volume of instruction in Russian, which was progressively replacing the national languages in non-Russian schools, and a continuous growth in the number of institutions where Russian was the sole operative language. Party rhetoric proclaiming the importance of Russian increased. The official view of a single Soviet ethnic group, the result of 'the convergence and fusion of peoples' (*sblizhenie i slijanie narodov*), was becoming a reality. The sphere of Russian usage spread beyond education to many administrative levels, including local-level administration. It had become the lingua franca of the USSR.

Mikhail Gorbachev came to power in 1985, and his restructuring program, *perestroika*, focused governmental attention on economic and political problems. Until 1989, both language and nationality policies remained essentially unchanged from previous years; in fact, they received little attention from the central government in Moscow. In many regions, the linguistic situation was essentially stable throughout this period; yet, a number of Republics began to change language policies within their own territories. By the end of the Soviet era, most of the territories that remained in the Russian Federation had adapted legislation to place the titular language, the language of the ethnic majority of their region, on an equal level with Russian.

References

Comrie, Bernard. 1981. The languages of the Soviet Union. Cambridge: Cambridge University Press.

Comrie, Bernard, Gerald Stone, and Maria Polinsky. 1996. The Russian language in the 20th century. Oxford: Clarendon Press.

Kirkwood, Michael. (ed.) 1989. Language planning in the Soviet Union. London: Macmillan.

Kreindler, Isabelle. (ed.) 1985. Sociolinguistic perspectives on Soviet national languages: their past, present and future. Berlin: Mouton de Gruyter.

Neroznak, V.P. (ed.) 1994. Krasnaia kniga iazykov narodov Rossii (Red book of the languages of the peoples of Russia). Moscow: Academiia.

Smith, Graham (ed.) 1990. The nationalities question in the Soviet Union. London: Longman.

LENORE A. GRENOBLE

Soviet Union: Successor States

Geographically, the former Soviet Union occupied one sixth of the Earth's land surface stretched from the Baltic Sea in the West to the Pacific Ocean in the East, from the Arctic Ocean in the North to the Black Sea and Tian Shan mountains in the South. It accommodated a number of ethnic groups with more than 100 languages, which can be combined into six major language families. There were 15 Union Republics with a total population of more than 240 million people, who shared common Soviet political experience for about 70 years.

The Soviet multilingual state originated from the state of Kievan Rus (ninth–twelfth centuries CE). By the fourteenth century, Moscow accumulated enough power to become a new political center of Slavic-speaking people. The expansion of Russia to the east, north, and south to the areas populated with non-Slavic people started in the sixteenth century. Russia annexed the lands of Kazan and Astrakhan Khaganats (Turkic-speaking people), defeating the Mongol Horde. They also conquered Finns in the northwest (Finno-Ugric language family), and proceeded further to the east, adding Siberia, the Far East, and Alaska to the Empire.

Russian Empire. By the end of the nineteenth century, the Russian Empire celebrated its last acquisition, the territory of Caucasus and Turkistan (present Central Asia). To run such a huge multiethnic state would have been impossible without promoting the use of a single language in all spheres of life including administration, legislation, education, etc.; hence, the Tsarist government began implementing its policy of russification. The policy implied a ban or restrictions on the use of languages other than Russian in government, education, the press, and books. In order to enforce this policy, the government encouraged the migration of ethnic

Russian population into other parts of the Empire and elevated the literacy rate through the study of Russian. The policy of russification was implemented by the Tsarist government in all of the provinces of the Russian Empire since the 1860s. However, in different parts of the Empire, the implementation of this policy was not uniform. On the one hand, in Ukraine any publications in Ukrainian language were prohibited, and the only language of instruction in Ukrainian schools and universities was Russian. Moreover, the Ukrainian and Belarusian languages were considered to be variations of the Russian language. On the other hand, people in Turkistan were not restricted in the use of their languages. Also, the Tsarist government did not attempt to change the Arabic-based alphabet of the Turkistanese into Cyrillic script. Turkistan enjoyed this freedom as a result of the Tsarist government's policy of avoiding unrest in newly acquired territory populated by people of different faith. However, the government encouraged Russians to move to Turkistan, thus providing a base for cultural and linguistic changes. By 1917, most of the major languages of the former Russian Empire had their own alphabets: for example, Central Asians and Azeri people used Arabic script, Baltic peoples used the Latin alphabet, and Armenians and Georgians had alphabets of their own.

Soviet Union. During the time of the Union of Soviet Socialist Republics or USSR (1922–1991), the language policy was a complex issue, revised with every change of the Soviet leadership. The Russian language preserved its dominating role and continued to serve as the lingua franca, enabling communication across the country's numerous ethnic groups. However, the Soviet government imposed several changes in language policy. In the mid-1920s, the Soviet government began a campaign to develop alphabets based on Latin script for most non-Slavic groups of the country. The Arabic script of the Muslim population was changed into Latin script. However, the strong and established traditions of written language in Georgia and Armenia withstood the government's campaign, and the ancient alphabets were preserved. The underlying rationale in the change of alphabets was the attempt of the Soviet government to cut off all the ethnic groups from their cultural, ethnic, religious, and ideological roots, and to begin implementation of the totally new culture and ideology of the Soviet state. Also, alphabets were created for those small ethnic groups who had never had a written language before.

In the late 1930s, there was yet another sudden change in the Soviet language policy, which was only possible due to tough political measures used by the Soviet government. The change, initiated under the rule of Stalin, was aimed at achieving the huge task of

unifying the scripts into Cyrillic, in order to strengthen the Soviet identity and begin the creation of the Soviet nation. The Latin script created for different ethnic groups in the 1920s and 1930s was replaced with a modified Cyrillic. Extra letters were added to the alphabet to reflect those sounds in non-Slavic languages that were not present in Russian.

At the same time, the Soviet government encouraged development of the literary language of each ethnic group, and each ethnic group could proudly name national writers who wrote in their language. Each Union or autonomous republic had radio and later TV channels broadcasting in their national language. Books, newspapers, and magazines were published in national languages as well. Nevertheless, the aim of the Soviet government to strengthen the role of the Russian language was never abandoned. Official propaganda put the Russian language as the language of Lenin, Stalin and the revolution, the language of the largest and greatest nationality, and the language of great Russian writers.

At primary and secondary education levels, people generally had a choice of the language of instruction since there were classes taught in Russian or in the language of the eponymous nationality, Kyrgyz in the Kirgiz SSR, for example, or Moldovan in the Moldavian SSR. At tertiary institutions, Russian was the major language of instruction, with some exceptions. For example, Ukrainian was used as the language of instruction in social sciences, art, and literature in Kiev, or Tajik was used in social sciences, art, and literature in Tajikistan. Similar arrangements existed in each of the 15 Union Republics. All official government business was conducted in Russian, with the exception of local governments in the Union Republics that used local languages as well as Russian. However, military officers gave their orders in Russian. The policy of russification that was continued by the Soviet government created a single literary language, that was used by most of the people in the USSR.

The government encouraged ethnic Russians to move to other parts of the country in order to promote cultural convergence. However, the first wave of Russian migrants did not go willingly. During the collectivization campaign of the 1920s and 1930s, millions of wealthy peasants had their property confiscated and were banished to Siberia, the Far East, or Central Asia, thus diluting the local population. Another wave of migration took place during and after World War II, when plants and factories were moved, together with their workers, from potential war zones in the European part of the USSR to Siberia and Central Asia. The last major wave of migration occurred in the 1950s during the virgin land campaign, when young people from all over the country came to

the Kazakh SSR with the aim of growing wheat in its vast wild steppe. After the completion of the campaign and establishment of kolkhozy (collective farms) in Kazakhstan, some of these people remained in Kazakhstan or moved to the neighboring republics.

During World War II, there was also a relocation of whole ethnic groups away from the war zones. Germans, Crimean Tatars, and Koreans were forcibly relocated to Central Asian Republics, as the Soviet authorities feared their collaboration with the occupation army.

This reshuffling of nations within the country led to a multinational composition of each republic of the Soviet Union. Often, children of at least five different ethnic groups could be sitting in one class. By 1991, the year of dissolution of the USSR, around 25 million of Russians lived outside Russia.

Post-Soviet Era. After the dissolution of the Soviet Union in 1991, 15 new countries appeared on the world map. This event was indicative of many dramatic changes in the lives of the peoples. Along with major economic, political, and social transformations, there were changes in the language policies in each of the new countries.

One of the prominent changes was that the status of the Russian language in all of the new countries, with the exception of the Russian Federation, was reversed. From its dominating position, it now became the language of Russian-speaking minorities. The languages of the titular nations of each new post-Soviet country were declared the state languages, and they are increasingly replacing Russian in all spheres of life, including education, mass media, and administration. Under the Soviet policy of russification, the Russians did not have to learn languages of titular nations of other Soviet republics. In Estonia, for example, according to the 1989 census, around 80% of ethnic Russians did not speak Estonian at all. Dramatic changes in the language policies in the newly independent republics forced a sizeable proportion of the Russian-speaking population (i.e. representatives of nontitular ethnic groups) to leave the places they were born in or lived in for decades and move to Russia, although a considerable number of Russians still live outside the Russian Federation. Here are some figures for 1998: 34.7% of the total population of Kazakhstan were Russians; 30.4% in Latvia; 28.1% in Estonia; 22% in Ukraine, 18% in Kyrgyzstan, 13.2% in Belarus; 13% in Moldova; 8.7% in Lithuania; 6.7% in Turkmenistan, 6.3% in Georgia; 5.5% in Uzbekistan; 3.5% (this number is declining due to emigration) in Tajikistan, 3% in Armenia, and 2.5% in Azerbaijan.

An outstanding example is that of the Baltic countries, Estonia, Latvia, and Lithuania. The Russians who arrived or were born in those countries after the 1940s were not legally considered citizens unless they passed a language examination. The Laws on Citizenship, adopted in the Baltic countries after independence, made most of the representatives of the Russian-speaking minority (Belarusians, Russians, and Ukrainians) 'noncitizens'. This policy created ethnic tensions between the titular nations and Russian-speaking minorities. For example, in Estonia the population of Russian-dominated Narva demanded secession or at least regional autonomy.

A unique language policy is observed in Kyrgyzstan. Although the Kyrgyz language was declared the state language, the Russian language was given official status in 1996 and became the state language along with Kyrgyz in 2000. The reasons were continuing ties with Russia, financial considerations, and the disappearance of the threat of russification, since the majority of the Russian-speaking population had left the republic.

Another change in language policy was demonstrated in the attempts of some of the newly independent countries to change the Cyrillic alphabet to either Latin or Arabic. The changes of the scripts were aimed at highlighting the idea of independence from Russia, of national state-building, and as a symbol of the revival of the nation's culture, traditions, and language. Financial considerations and the enormous complexity of this task prolonged the change of the script in some of the republics. Other republics postponed it indefinitely. In Turkmenistan, a Latin-based Turkish alphabet replaced Cyrillic. Moldova switched to a Latin-based script. Tajikistan is gradually implementing a change to Arabic. In Uzbekistan, the script has been changed to Latin. In Azerbaijan, the alphabet reform of 2001 implies transition to the Latin alphabet. Kyrgyzstan retains the Cyrillic script.

In the Russian Federation, the language situation also changed after 1991. The Russian language is still the lingua franca of this large multinational state. However, the uncontrollable inflow of the foreign, mainly English-based, words in mass media and colloquial language since the cultural isolation of Russia from the West ended together with the Soviet Union. This trend changed the Russian language dramatically. The literary language created during the Soviet time is deteriorating, and new language standards are being created.

Language Families

The 15 former Soviet Union republics bring together six language families: Indo-European, Altaic, Uralic, North-Western, North-Eastern, and South-Caucasian. There are also some limited language groups and separate languages in the far eastern Siberia, which can

FAMILY	Group	Main Languages
Indo-European		
	East-Slavic	Russian, Belarusian, Ukrainian
	Baltic	languages Latvian, Lithuanian
	Romance	languages Moldavian/Romanian
	Iranian	language Tajik, Kurd, Osetin
		languages Armenian language
Altaic		
	Turkic	Azeri, Kazakh, Kyrgyz, Turkmen, Tajik,
		Tatar, Uzbek and other languages
	Manchu-Tungus	Evenk, Even languages
	Mongolian	Buryat, Kalmyk languages
Uralic		
	Finno-Ugric	Estonian, Finnish, Khanty, Mansi, Mari,
		Mordvin, Karelian, Udmurt,
		Komi, Komi-Permyak, Veps languages
	Samoyedic	Nenets, Selkap languages
North-Western Caucasian		Adyghey, Kabardyn, Abkhaz, Circassian languages
South-Eestern Caucasian		Chechen, Lezginian languages
South Caucasian		Georgian language
Paleoasiatic		Chukcha, Koryak, Kamchadal languages and others

Figure 1. Language Families and Groups.

be grouped as a Paleoasiatic group. The languages of this group are not related and are placed together only geographically.

The Indo-European Family

East Slavic Languages: The Russian Federation, The Ukraine, Belarus

Russian. The largest successor to the Soviet Union is the Russian Federation. Its total population in 1998 was 146,001,176, which was represented ethnically as follows: 81.5% of Russians, 3.8% of Tatars, 3% of Ukrainians, 1.2% of Chuvashs, 0.9% of Bashkirs, 0.8% of Belarusians, 0.7% of Mordvins, and 8.1% of others. The Russian Federation is home to more than 70 distinct ethnic groups, some of which are extremely small and the six mentioned above are more than a million of people in each. There are 21 autonomous republics in the Russian Federation as well as ten autonomous districts and one autonomous region.

In spite of enormous ethnic diversity, the official language of the Russian Federation is Russian. The language is also spoken as a secondary language in other republics of the former Soviet Union as well as in some countries of Eastern Europe (former socialist block). Three dialects of Russian, i.e. northern, southern, and central, are only distinct from each other by minor variations in the pronunciation of some consonants and vowels, unlike the significant differences in dialects of European languages. Literary Russian is based on the Moscow (central) dialect and is written in Cyrillic.

Ukrainian. The second largest Slavic state of the former Soviet Union is the Ukraine. In 1998, its population was 49,153,027 with the following ethnic composition: 73% Ukrainians, 22% Russians, 1% Jewish, and 4% other ethnic groups. The Ukrainian language is the state language; Russian and Polish are also spoken. The Ukrainian language is also spoken in Ukrainian communities in Belarus, Russia, Poland, and Slovakia.

Ukrainian is the direct descendant of the language spoken in Kievan Rus and is written in the Cyrillic alphabet. Distinct differences between Russian, Belarusian, and Ukrainian languages emerged somewhere between the twelfth and thirteenth centuries; however, development of the standard Ukrainian

language was postponed due to long political subordination to Russia and Imperial Russia's policy of restrictions in usage of the Ukrainian language. Modern literary Ukrainian language began its development only at the end of the eighteenth century.

Belarusian. The third branch of east Slavic languages is spoken in Belarus (or White Russia). In 1998, there were 10,366,719 people and 77.9% of them were Belarusians. There were also 13.2% Russians, 4.1% Polish, 2.9% Ukrainians, and 1.9% others. The Belarusian language used to be the official language of Lithuania in the thirteenth–sixteenth centuries when Belarus was a part of Lithuania. After the merging of Poland and Lithuania in the sixteenth century, Belarus found itself under Polish jurisdiction until the eighteenth century. There are a large number of borrowings from the Polish language. Belarusian became the official language of Belarus from 1990. Russian language became the second official language of Belarus in 1996. Belarusian is written in the Cyrillic alphabet.

Baltic Languages: Latvia, Lithuania

Latvian. Is the official language of Latvia. Out of the total population of Latvia in 1998 (2,404,926), there were 56.5% Latvians, 30.4% Russians, 4.3% Belarusians, 2.8% Ukrainians, 2.6% Polish, and 3.4% others. In 1989, Latvian language replaced Russian in administration, education, media, etc., although Russian and some other languages are still used in everyday life. Latvian is written in the Latin alphabet.

Lithuanian. Is the official language of Lithuania. In 1998, there were 80.6% Lithuanians out of the total population of the country of 3,620,756 people, 8.7% Russians, 7% Polish, 1.6% Belarusians, and 2.1% others. Along with the official language, Polish and Russian are spoken as well. The first written documents in Lithuanian are recorded back to the sixteenth century. The Lithuanian language is written in the Latin alphabet.

Romance Languages: Moldova

Moldavian/Romanian. The only country of the former Soviet Union where people speak one of the Romance languages is Moldova. The population of Moldova consisted of 4,430,654 people in 1989 and constituted 64.5% Moldovans, 13.8% Ukrainians, 13% Russians, 3.5% Gagauz, 2% Bulgarians, 1.5% Jewish, and 1.7% others. In 1989, the Moldavian language (virtually the same as Romanian) was declared the official language; however, in 1994 the enforcement of the Law on Languages was relaxed due to the

rise of separatist tendencies from Slav-populated areas of Dnestr basin. The Cyrillic alphabet was replaced with Latin.

Armenian: Armenia

Armenian is spoken by a wide Armenian population worldwide. It is the language of Armenians in Armenia, Turkey, and other parts of the former Soviet Union, Middle East, America, Romania, Poland, France, and others. In 1989, in Armenia itself, there were 3,344,336 people and 93% of them were Armenians, 3% were Azeri (nearly all Azeri emigrated in 1993), 2% were Russians, and 2% were others. It is traditionally believed that the Armenian alphabet was created by the scholar and bishop St. Mesrop on the basis of Greek and Aramaic scripts sometime around 405–406 CE. In the twelfth century, there were two more letters added to the alphabet. Armenian literature appeared by the fifth century CE and the modern literary Armenian started developing in the nineteenth century.

Iranian Languages: Tajikistan, The Russian Federation, Georgia, Armenia, and Azerbaijan

Tajik. Is spoken mainly in Tajikistan. There were 6,440,732 people in Tajikistan in 1998 with the following ethnic composition: 64.9% Tajiks, 25% Uzbeks, 3.5% Russians (the number is declining due to emigration), and 6.6% others. Tajik language has been the official language of the country since 1989, although Russian is used in government and business. There are a large number of borrowings from Arab, Uzbek, and Russian languages. Before 1930, it was written in Arabic, when the script was changed to Latin; in 1940, it was changed to Cyrillic. The postindependence government is gradually implementing a change to Arabic.

Osetin and Kurd. There are also other small groups of Iranian-speaking people. Osetins are dispersed throughout Caucasus (Georgia, the Russian Federation). The first written document in the Osetin language is dated back to 941 CE on the basis of the Greek alphabet. In 1844, Shegren created the alphabet on a Cyrillic base, which was changed to Latin in the 1920s. In 1938, the Northern Osetian alphabet was changed to Cyrillic and the South Osetian alphabet was changed to Georgian. The latter was changed to Cyrillic in 1954.

Kurds populate areas in Turkmenistan, Armenia, Georgia, and Azerbaijan, Turkey, Afghanistan, and Lebanon. The Kurds of the Russian Empire used the Arabic alphabet; however, in 1921, it the Soviet government changed it to Armenian, in 1929 to Latin, and in 1946 to Cyrillic.

Altaic Group

Turkic Languages: Azerbaijan, Kazakhstan, Kyrgyzstan, Moldova, Turkmenistan, The Russian Federation, Uzbekistan

Turkic-speaking peoples live mainly in the Central Asian Republics, although there are some Turkic groups in the middle Volga region of Russia (Bashkirs, Tatars, Chuvash) and in the Caucasus (Azeri, Balkars, Karachays, Kumyks, and Nogays). In Siberia, there are numerous groups of Turkic-speaking population who live between the Ural and Lake Baikal (Altais, Khakass, Tofalars, Shors, Tuvans) and also Yakuts who mainly live in the middle Lena basin and Dolgans in the Arctic.

Azeri. Azeri is the state language of Azerbaijan. The total population of the country in 1998 was 7,748,163 people, 90% of whom were Azeri, 3.2% Dagestani, 2.5% Russians, 2% Armenians, and 2.3% were represented by other ethnic groups. The Russian language is still used in the country.

The literary Azeri language began its development from the eleventh century. Modern Azeri is based on Baku and Shemakhan dialects. The original Arabic script used by Azerbaijani people was replaced by Latin in 1929 and by Cyrillic in 1939. After 1991, the government introduced a language reform, part of which was the change to Latin script.

Kazakh. Is mainly spoken in Kazakhstan, which has the second largest population, after Uzbekistan, of the former Soviet Central Asia (16,733,227 people (1996)). The Kazakh language is also spoken in Sinkiang Uygur autonomous region in China and in Uzbekistan, Mongolia, and Afghanistan. In 1996 there were 46% Kazakhs, 34.7% Russians, 4.9% Ukrainians, 3.1% Germans, 2.3% Uzbeks, 1.9% Tatars, and 7.1% others in Kazakhstan. Kazakh people used the Arabic alphabet, which was changed into Latin and later to Cyrillic. The Kazakh language is the state language of Kazakhstan.

Uzbek. Uzbek is spoken primarily in Uzbekistan and in some parts of Kazakhstan, Tajikistan, Turkmenistan, Afghanistan, and China. In 1996, there were 49,153,027 people in the country; 80% of them were Uzbeks, 5.5% Russians, 5% Tajiks, 3% Kazakhs, 2.5% Karakalpaks, 1.5% Tatars, and 2.5% others. The Uzbek language is the official language of the country; however, Russian and Tajik are used as well. Although the Arabic script was used for written language before the Bolshevik Revolution, it was changed first to Latin in 1927 and to Cyrillic in 1940. The postindependence tendency was to change Cyrillic into Latin.

Kyrgyz. Is the state language of Kyrgyzstan with a population of 4,685,230 in 1996. A little more than half the population were Kyrgyzs (52.4%), 18% Russians, 12.9% Uzbeks, 2.5% Ukrainians, 2.4% Germans, and 11.8% other ethnic groups. The Kyrgyz and Russian languages are the state languages. Kyrgyz language used Arabic alphabet up to 1926 when it was changed to Latin; it was changed to Cyrillic in 1940.

Turkmen. Turkmenistan is a country of 4,518,268 (July 2000 est.) people. In 1995, there were 77% Turkmens, 9.2% Uzbeks, 6.7% Russians, 2% Kazakhs, and 5.1% others. The languages spoken are Turkmen, Russian, and Uzbek. The Turkmen language became the official language of Turkmenistan in 1990. A Latin-based Turkish alphabet was introduced in 1993, replacing Cyrillic.

Manchu-Tungus Languages: The Russian Federation

Another branch of Altaic languages, Manchu-Tungus languages, is spoken by the Evenks, Evens, and other small groups dispersed throughout eastern Siberia. These languages had never been written before the Russian revolution; presently, they are written in Cyrillic.

Mongolian Languages: The Russian Federation

The Mongolian languages are spoken in the Lake Baikal region (Buryats) and to the west of the lower Volga (Kalmyks). These languages got their first alphabet, Cyrillic, after the Russian revolution.

The Uralic Group

Finno-Ugric Languages: Estonia, The Russian Federation

The Finno-Ugric languages are spoken in Estonia and the European part of the Russian Federation from northwest (Karelians, Finns, and Veps) to the upper Volga river and the Ural mountains (the Mari, Mordvins, Udmurt, Komi, and Komi-Permyak) to lower Ob' river basin in Siberia (the Mansi and Khanty). The areas of Finno-Ugric-speaking population are intertwined with areas of other language families such as Slavic, Turkic, Baltic, and others. Close contacts with other people led to a large number of borrowings from Russian, Turkic, Romance, Iranian, German, and Baltic languages. Practically all of the Finno-Ugric languages, except Estonian, use the Cyrillic script.

Estonian. The total population of Estonia was 1,431,471 people in 1998. Out of them, Estonians constituted 65.1%, Russians 28.1%, Ukrainians 2.5%,

Belarusians 1.5%, Finns 1%, and others 1.8%. Estonians are the largest group of Finno-Ugric people in the former Soviet Union. Out of two main dialects, northern and southern, the former (or Tallinn) dialect served as a base for the Estonian literary language. The language is most closely related to Finnish, Livonian, Ingrian, Karelian, and Veps languages. It is written in Latin.

The Samoyedic Languages: The Russian Federation

The Samoyedic group is also quite dispersed in the Siberian tundra and taiga from Kola Peninsula to the Yenisey river (spoken by Nenets), around the middle Ob' river (spoken by Selkup) and in Taymyr Peninsula (spoken by Nhanasan). Samoyedic languages are agglutinative in structure and are spoken by a small number of people.

The Caucasian Family

This group is made up of more than 40 languages, that are spoken by about six million people. There are three main branches: North-Western (Abkhazo-Adyghian), North-Eastern (Nakho-Dagestanian), and South (Kartvelian).

North-Western: The Russian Federation, Georgia (Abkhazia)

Adyghey, Kabardyn, Abkhaz. It is characteristic of these languages that they obtained alphabets after the Russian revolution. Adyghey language was first written in Arabic script, which was later changed to Latin (1927) and then to Cyrillic (in 1938). The Abkhazian language was written in Latin since 1928; the script was changed to Georgian in 1938, and then to Cyrillic in 1954. Kabardin language was also first written in Latin (from the mid-1920s), and from 1936 in Cyrillic.

North-Eastern Languages: The Russian Federation, Dagestan, Azerbaijan

Chechen is spoken in Chechnya and Dagestan (the autonomous republics in the Russian Federation). The original Arabic script was changed to Latin in 1927 and to Cyrillic in 1938.

Lezginian language is spoken in Dagestan and Azerbaijan. The Arabic script was used for writing before the Russian revolution; in 1928, it was changed into Latin and in 1938 to Cyrillic.

South Caucasian Languages: Georgia

Georgian. Georgia's population in 1998 reached 5,019,538 people. There were 70.1% Georgians, 8.1% Armenians, 6.3% Russians, 5.7% Azeri, 3% Osetians, 1.8% Abkhaz, and 5% others. Georgian is the official language, although Russian, Armenian, and Azeri are used as well.

Georgian is written in its own alphabet, which is believed to have been derived by King Parnavaz in the third century BCE on the basis of the Aramaic script. The script was subjected to revision under the influence of Greek alphabet when Georgians embraced Christianity in the fourth century CE. In 1926, the Soviet government attempted to replace the alphabet with Latin with no success. Georgian is the only south Caucasian language with old literary traditions. Other south Caucasian languages, such as Mingrelian, Swan, and Laz, acquired written scripts and developed their literary languages mainly after the Russian Revolution.

Paleoasiatic Languages

At the outer edge of far eastern Siberia, there are numerous ethnic groups that speak different languages. One group is the Chukcha (Luorawetlan), Koryak (Nymylan), and Kamchadal (Itelmen), which are spoken from Chukotka to Kamchatka. Another group is the Eskimo-Aleut group. There are also completely isolated languages of people inhabiting Sakhalin Island, lower Amur river, Kolyma lowland, and middle Yenisey river. Some of these languages were influenced by neighboring Yakuts, who speak a Turkic language, and all of them borrowed considerably from Russian. The written scripts of paleoasiatic languages were developed after the Russian revolution.

References

Baskakov, Nikolai A. 1953. The Turkic language of Central Asia: problems of planned culture contact. London: Central Asian Research Centre.

Baskakov, Nikolai A. (ed.) 1982. Razvitie natsionalnykh iazykov v sviazi s ikh funktsionirovaniem v sfere vysshego obrazovaniia (Development of national languages in relation to their functioning in higher education). Moskva: 'Nauka'.

Bol'shaya sovetskaya entsyklopedia, edited by Prokhorov, A.M. Moscow: Sovetskaya Entsyklopedia, 3rd edition, 1973 (Great Soviet encyclopedia, New York: Macmillan, Inc., 1983).

Chinn, Jeff, and Robert Kaiser. 1996. Russians as the New Minority: ethnicity and nationalism in the Soviet Successor States. Boulder, CO: Westview Press.

Comrie, Bernard, Gerald Stone, and Maria Polinsky. 1977. The Russian language since the revolution. Oxford: Clarendon Press; as The Russian language in the twentieth century, 2nd edition, Oxford: Clarendon Press, 1996.

Huttenbach, Henry R. 1990. Soviet nationalities policies: ruling ethnic groups in the USSR. London, New York: Mansell.

Kirkwood, Michael (ed.) 1989. Language planning in the Soviet Union. London: Macmillan.

Laitin, David D. 1998. Identity in formation: the Russian-speaking populations in the near abroad. Ithaca and London: Cornell University Press.

Lewis, Glyn E. 1972. Multilingualism in the Soviet Union: aspects of language policy and its implementation. The Hague: Mouton.

Magner, Thomas F., and William R. Schmalstieg (eds.) 1970. Baltic linguistics. University Park: Pennsylvania State University Press.

Matthews, William K. 1951. Languages of the U.S.S.R. Cambridge: The University Press.

Menges, K.H. 1995. The Turkic languages and peoples: an introduction to Turkic studies, 2nd, revised edition. Wiesbaden: Harrassowitz.

Smith, Graham (ed.) 1996. The Baltic states: the national self-determination of Estonia, Latvia, and Lithuania. Basingstoke: Macmillan.

ALFIA ABAZOVA

Spain

Spain, a country on the Iberian Peninsula in Western Europe, occupies 504,782 square kilometers (194,897 square miles) and has a population of about 40 million people. It is home to Castilian Spanish, a language that has left its linguistic mark in former colonies throughout Latin America and the Philippines, as well as large Spanish-speaking communities in a number of other countries. There are, however, many other languages that also claim Spain as their home; these include historical languages (Iberian, Celtiberian, and Mozarabic), other Romance languages (Catalan, Galician, Aranese, Aragonese, and Asturian), and one language isolate (Basque).

Historical (Extinct) Languages

Several inscriptions dating from the fifth to the first centuries BCE have been found in eastern Spain in a language known as Iberian. Most of these inscriptions are written in the Iberian script, but a small number were written in the Greek script. Scholars have unsuccessfully attempted to find a relationship between this unknown non-Indo-European language and Basque, a language isolate spoken in northern Spain.

Celtiberian is an extinct continental Celtic language (as opposed to the more familiar insular Celtic languages such as Gaelic, Irish, Welsh, etc.) that was spoken in northeastern Spain. The Celtiberians may have inhabited Spain as early as eight BCE. Celtiberian inscriptions and texts date between early two BCE to around one CE. There was no Celtiberian script; earlier inscriptions are written in the Iberian script, and some of the later texts are written in the Latin alphabet. These texts and inscriptions provide important grammatical information but little by way of vocabulary as they tend to be short.

Introduction of Latin

At the beginning of the second century BCE, the Romans entered the Iberian Peninsula and brought with them their language, Latin. This new language gradually absorbed all of the languages, with the exception of Basque, spoken on the peninsula. As a result, all languages currently spoken on the Iberian Peninsula, except Basque, are Romance languages. It is assumed that the original languages of Spain (Iberian, Celtiberian, and Basque) influenced the Latin spoken on the peninsula, contributing to the characteristics of the various vernaculars. A dialect continuum developed on the peninsula; over time, the Latin of the various parts began to differ more and more from one another. By the eighth century CE, five distinct language groups appeared: Galician-Portuguese, Asturian-Leonese, Castilian, Aragonese, and Catalan. Three particular areas in the north gradually gained certain social prestige, thus establishing their eventual domination: Galician-Portuguese, based around Santiago de Compostela in the northwest; Castilian, around Burgos in the north; and Catalan, around Barcelona in the northeast.

A now-extinct Arabic-influenced Ibero-Romance language, Mozarabic, developed in the south during the Middle Ages and was spoken by Mozarabs (Christians under Moorish rule), Muslims, and Jews. No standard language developed because Mozarabic was rarely written; when it was written, it was written in either the Arabic or Hebrew script. Information on this language comes down to us in the following forms: poems written in the Arabic or Hebrew script that form part of longer compositions in Arabic or Hebrew; words and word lists written in the Arabic script; place names; some medieval and early modern Arabic-Latin and Arabic-Spanish dictionaries; and certain post-Reconquest legal documents. Slowly, Portuguese, Castilian, and Catalan speakers reconquered the territories where Mozarabic was spoken. Mozarabic was probably mutually comprehensible with these varieties of Romance but was gradually absorbed by the new dominant language until its disappearance as a separately identifiable language by 13 CE. Some have argued that Mozarabic influenced the

other languages as it was being absorbed, but this is only apparent in vocabulary.

Modern Languages

Because of its instrumental position during the Reconquest, Castile became a dominant force on the peninsula. Castile became politically superior to the other regions, and Castilian, the language of the court, government, and the expanding empire, gained prestige as a result; by the fifteenth century CE, its use as the language of culture and administration spread to non-Castilian areas. In some areas like Galicia, Castilian became the language of the elite, giving rise to diglossia—a situation wherein a speech community uses two languages (or two very distinct varieties of one language) for different social contexts or for performing different functions.

Attempts to use language policies to strengthen Spanish national identity began to be considered in the seventeenth century CE. In addition, three important institutions furthered the position of Castilian throughout Spain: the Catholic Church, the education system, and the military. Charles III declared in 1768 that Castilian was to be used as the language of administration and education throughout Spain. The Castilianization of Spain continued until the latter half of the nineteenth century when a resurgence of cultural activities, including codification and elaboration, in non-Castilian languages occurred; the Romantic Movement motivated these 'renaissances'. In 1932, the king abdicated, and the four-year Second Republic was proclaimed. Greater powers in the form of statutes of local autonomy were given to Catalonia, Galicia, and the Basque Country during this time, although the Civil War meant that, with the exception of Catalonia, they were not able to experience this autonomy properly.

With the end of the Civil War came a dictatorship (Franco's regime), which repressed regional languages to centralize the country. Speaking non-Castilian languages was viewed as antipatriotic and was considered rebellious; punishments included fines and imprisonment. In 1966, the Freedom of Expression Law was passed, which removed strict censorship and allowed instruction in non-Castilian mother-tongue languages and publishing in these languages.

Franco died in 1975, and three years later, the 1978 Constitution was passed. This turned Spain into a Western-style democracy and resulted in the legal recognition of the other languages. Castilian was still the official language of the country, but each of the autonomous communities could declare their own official language alongside Castilian.

Catalan (*català*) (called Valencian (*valencià*) in Valencia) is spoken by approximately 6.5 million people in the autonomous communities of Catalonia, Valencia, the Balearic Islands, and Aragon. From the eleventh to the fourteenth centuries, Catalan was considered a prestigious language comparable to French, Italian, Castilian, and Portuguese. Literature and philosophical works thrived as Catalan replaced Latin and Provençal as the language of cultural and literary production. From the sixteenth century, however, Catalan declined in prominence as Castile's power and control increased. Catalan lost its prestige during this time and largely became a spoken language. This persecution and Castilianization continued until the *Renaixença* 'Renaissance' of the second half of the nineteenth century. Catalan has perhaps had the most successful revival of any of the other minority languages; this may be due to the fact that Catalan had an extensive literary heritage that was written, in contrast to Galician which had more of an oral literary heritage and Basque which did not have much of a literary heritage. In 1914, the *Institut d'Estudis Catalans* 'Institute of Catalan Studies' was set up by the *Mancomunitat*, the predecessor of the current *Generalitat*, the governing body of Catalonia.

Galician (*galego*) is spoken by approximately 2.3 million people in northwestern Spain in the autonomous region of Galicia, where it is one of the official languages. Galician was an important literary language during the Middle Ages and gained prestige as a result, especially in lyric poetry; speakers of Castilian and Leonese frequently composed poetry in Galician because of this prestige. Over the course of the fourteenth century, however, Galician lost ground to Castilian due to the large influx of Castilian nobles and Castilian-speaking priests. This relegated Galician to the low position of a popular spoken form used by peasants. Galician was maintained by the large number of Galician speakers who immigrated to other parts of the Spanish empire, particularly the Americas. Galician did not revive until its *Rexordimento* of the second half of the nineteenth century; the Academy of the Galician Language was founded in 1906.

Basque (*euskara*) is currently spoken by about 565,000 people in the Autonomous Basque Community and the Autonomous Community of Navarre. Not only is it the oldest language still spoken on the Iberian Peninsula but it is also the only non-Indo-European language; it is a language isolate—its relationship with other languages has not yet been (nor, perhaps, will ever be) established. Basque speakers accepted Castilian dominance more readily than the Catalan speakers because, in exchange, they were able to retain their rights. Unlike Catalan and Galician speakers, Basque speakers did not have a large body of written literature to unite their language. The Basque language divided into several dialects, some of which are no longer mutually intelligible.

Aranese (*aranès*) is spoken by approximately 6,000 people in the Val d'Aran, a Pyrenean valley on the French–Spanish border forming a part of the Catalonia Autonomous Community. It is a dialect of Gascon, a variant of Occitan, spoken in southwestern France and is, along with Catalan and Castilian, an official language in the Val d'Aran. In April 1983, Catalonia enacted the Linguistic Normalization Law, as a result of which five steps were taken in support of maintaining Aranese: Aranese was declared an official language of the Val d'Aran, and the right to know and use this language in relations and in public ceremonies within the territory was recognized; the *Generalitat* of Catalonia (the governing body of Catalonia), along with institutions in the Val d'Aran, became obligated to do what was necessary to guarantee the knowledge and normal use of Aranese and to encourage its normalization; the Aranese form of place names in the Val d'Aran was declared the official form; the *Generalitat* became obligated to supply the means necessary to guarantee the teaching and use of Aranese in the schools of the Val d'Aran; and the *Generalitat* was required to take the necessary steps to ensure that Aranese is used in the mass media of the Val d'Aran. The immediate consequences of this law include the creation of a commission to standardize Aranese, the publication of textbooks in Aranese for their use in schools, and a sociolinguistic survey to provide a basis for possible future actions. One of the most important steps, however, was the introduction of Aranese into schools. Aranese is the language of instruction during the first few years of education, although Castilian and Catalan take leading roles later. The Parliament of Catalonia reintroduced the General Council (*Conselh Generau*) into the Val d'Aran, which regularly uses Aranese, and certain government powers were attributed thereto. In addition, there is one program a week on the Catalan television service in the Aranese language and a short daily program in the radio schedule; one Catalan newspaper has a weekly Aranese supplement.

Aragonese is spoken in a few Pyrenean valleys by about 30,000 people in the Aragon Autonomous Community, where it is, along with Castilian and Catalan, an official language. Aragonese appears in some schools as an optional extracurricular subject. It is not used in television; however, it can be heard on the radio, and some infrequent magazines are published in Aragonese. There is a newspaper in North Aragon that includes a weekly Aragonese supplement.

Asturian is spoken in the Asturian Autonomous Community and the north and west parts of Castile-León, and although it is not recognized as an official language, its protection and promotion is stated in the *Statute of Autonomy of Asturias*. Asturian is present in some preprimary and primary schools as the language of instruction and is offered as an elective subject in primary and secondary schools. At the University of Oviedo and in teacher-training college, Asturian is offered as an option. Asturian television programs practically do not exist, although the parliament has approved the procedural stage for a project to create a public television and radio broadcasting institution in Asturias. There is also one weekly newspaper, *Les Noticies*, extracts in other newspapers, and a few magazines published in Asturian. However, the presence of Asturian in government is minimal; only a few departments accept documents written in Asturian, and it is rarely used in courts, legislative texts, or public signs.

References

Campbell, George L. 1991. Compendium of the world's languages. London and New York: Routledge, 2nd edition, 2000.

Lacombe, Georges. 1952. Langue Basque. Les Langues du Monde, ed. by A. Meillet and Marcel Cohen. Paris: Centre national de la recherche scientifique.

Mar-Molinero, Clare. 2000. The Iberian Peninsula: conflicting linguistic nationalisms. Language and nationalism in Europe, ed. by Stephen Barbour and Cathie Carmichael. Oxford and New York: Oxford University Press.

Price, Glanville. 1998. Encyclopedia of the languages of Europe. Oxford and Malden, MA: Blackwell Publishers.

Siguán, Miguel. 1992. España Plurilingüe. Madrid: Alianza Editorial; as Multilingual Spain, translated by Miguel Siguán. Amsterdam: Swets & Zeitlinger, 1993.

ROBERT A. CLOUTIER

See also **Basque; Diglossia; Indo-European 4: Romance; Spanish and Iberoromance Languages**

Spanish and Iberoromance Languages

Spanish and the Iberoromance languages derive from Latin, which was brought into the Iberic peninsula in the third century BCE, and superseded most of the previously spoken Celtic and non-Indo-European languages. At the time of the Germanic invasion, in the fifth century CE, the Iberic peninsula was not yet

fully romanized, as shown by the survival of the Basque language in the northern region, near the gulf of Guascogne, a language presumably spoken by a much larger part of the population in pre-Roman times. The Germanic people who poured into the area of the Roman empire left deep linguistic traces in the Iberic languages, much in the same way as in the other Romance languages. Peculiar of the Iberic area is the influence of Arabic, following the period of domination from the eighth century BCE to the 'Reconquista', fully accomplished only in 1492.

Drawing the linguistic limits of the Iberoromance area is not without problems. Castilian Spanish and related vernaculars, Portuguese, and Galician certainly belong to the Iberic branch of Romance, but the position of Catalan remains more controversial: both geographically and linguistically, Catalan was closer in origin to the Galloromance branch, especially to Occitan; only later cultural and historical developments created more solid links with the languages of the Iberic peninsula.

The Iberoromance languages share with the rest of Western Romance the palatalization of Latin /kt/, which became /it/, and the retention of -s plural. Other features of Iberoromance include the use of verbs derived from Latin *tenere* (in Portuguese) and *stare* as auxiliaries (the latter also found in the Southern Italian varieties excluding Sicilian) and the prepositional accusative, with the extension of the preposition *a* to animate definite direct objects, which, however, is nowadays fully standardized only in Spanish (note that the prepositional accusative is also found in most Southern Italian varieties, in Sardinian and in Rhetoromance; a similar development in Rumanian involves the preposition *pe*, from Latin *per*).

Spanish (Castilian)

With its more than 330 million speakers, Spanish is one of the most spoken languages in the world, second only to Chinese (data from the 1996 edn. of *Ethnologue*). It is the official language of Spain as well as of Argentina, Bolivia, Columbia, Costa Rica, Cuba, Chile, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panamá, Paraguay, Peru, Puerto Rico, Dominican Republic, Uruguay, and Venezuela, and it is spoken as mother tongue by more than 20 million speakers in the United States.

Given its diffusion, Spanish has different standards. Usually, a distinction is made between the peninsular standard, based on Castilian, and the American standard, whose prestige center is Mexico City; the most relevant deviations from this standard in Latin America are found in Argentina.

Special features of Spanish as a Romance language include diphthongization of Latin short /e/ and /o/ under stress in both closed and open syllables, the change /f/ > /h/ in word-initial position (/f/ is retained as such before /ue/), further palatalization of /kt/ (> /it/ in all Western Romance and > /tʃ/ only in Spanish), and lenition of voiced stops in the internal position. The vowel system includes five phonemes, /a/, /e/, /o/, /i/, and /u/, with no distinction between long and short vowels, and there are no cases of vowel reduction. The most important phonological feature that builds an inner isogloss within Spanish varieties is the existence of a distinction between /s/ and /θ/, typical of peninsular standard, but not found in Andalusian and American varieties (see below). Morphologically, vernacular varieties still spoken today are Leonés, in the province of Leon, which shares a number of features of Gallego and Portuguese, Argonés, in the province of Huesca, and Andalusian, much closer to Castilian, in Andalucia. This last variety has among one of its features the so-called *seseo*, that is, convergence of /θ/ and /s/ in /s/ (the opposite phenomenon, called *ceceo*, i.e. convergence of both phonemes in /θ/, is also sporadically attested). Since *seseo* is also found in all Latin American varieties, it was formerly assumed that the latter were based on Andalusian; however, there is evidence for the existence of two spirants in early Mexican Spanish. Most likely, the two spirants were articulatorily closer to each other than /s/ and /θ/, and eventually merged in both Latin America and Andalucia, but became more distinct in Castilian. So /θ/ is a Castilian innovation, which dates after the conquest of Mexico.

In the field of morphology, the largest dialectal variation is found in the pronominal system and the use of verb forms for second person plural. Standard peninsular Spanish has the following set of (accented) personal pronouns:

	Singular	Plural
1.	yo	nosotros/nosotras
2.	tú	vosotros/vosotras
3.	él/ella	ellos/ellas
Polite form	usted	ustedes

The polite form *usted/usetedes*, which, in the standard, agrees with third-person singular/plural verb forms, was created in the course of the Middle Ages; as in some other Romance varieties, most notably Italian, there has been a long oscillation between third-person forms and second-person plural. The form *vos* (an earlier second person plural) is still found in literary peninsular Spanish of the nineteenth century and lies at the basis of *voseo*, i.e. the substitution of second-person singular *tú* with *vos* (see below). The form *vosotros/vosotras* for familiar second-person plural only occurs in Castilian; the other varieties have *ustedes*, either followed by second-person plural verb forms (Andalusian,

Canarian) or by third-person plural verb forms (Latin America). In Argentina, where *voseo* is accepted at the level of the written standard, *vos* is used for familiar second-person singular, followed by verb forms derived from second-person plural, which no longer correspond to the forms used in the peninsular standard (*vos hablás*, 'you (sg.) speak'; the second-person plural of *hablar* in peninsular standard Spanish is *habláis*), and *ustedes* is used for second-person plural, both familiar and polite, followed by third-person verb forms. At the substandard level, *voseo* is found in all Latin America, with the exception of Mexico and Peru, the countries that maintained closer links with Spain after the conquest, due to the existence of the Viceroyalties.

Another peculiarity of Castilian in the field of pronouns is so-called *leísmo*. In the literary standard, *leísmo* is confined to the use of the clitic *le* instead of *lo* for masculine animate definite objects, thus matching (but only for the masculine singular) the distribution of prepositional accusative in the case of noun phrases (or accented pronouns), as in *Maria vi a Juan/Maria le vi* vs. *Maria vi el libro/Maria lo vi* (some speakers, especially in the Madrid and Valladolid areas, also use *le* for inanimate objects).

A variety of Spanish is Judeo-Spanish, or Sefardí, which derived from the language spoken in Spain at the time of the expulsion of Sephardic Jews, in 1492. The expelled Jews moved to different countries, mainly on the shores of the Mediterranean, where small communities thrived until the seventeenth century. Their language, which retains a number of archaic features (for example, retention of initial /f/), is still spoken today by small groups, mostly in Israel.

The first written attestations of Spanish date to the tenth century CE. Early poetic compositions in the Iberic peninsula were written in a southern vernacular, Mozarabic, the language spoken during the Arabic domination. The first literary text in Medieval Castilian is the epic *Cantar del mio Cid*, written around 1140, which narrates the deeds of the hero of the Reconquista.

Portuguese

Portuguese is the official language of Portugal, where it is spoken by about 10 million speakers, and of Brazil; according to the 1995 edition of the *World almanac*, the total speakers of Portuguese were 170 million. European and Brazilian Portuguese differ considerably from each other in various respects.

Portuguese vocalism is rather complicated, if compared with the system of neighboring Spanish, and more similar to the French system, on account of the existence of a phonemic opposition between oral and nasal vowels, and of extensive vowel reduction in unstressed syllables. Two different vowel systems can be set up for stressed and unstressed syllables. In stressed syllables, eight oral vowels and five nasal vowels are distinguished: /i/, /e/, /ɛ/, /ə/, /a/, /ɔ/, /o/, /u/ and /ĩ/, /ẽ/, /ã/, /õ/, /ũ/. In unstressed syllables, the number of distinctions is considerably reduced: in particular, final unstressed vowels are reduced to /ə/, /ɐ/, /u/ and non-final vowels to /i/, /u/, /ə/, /ɐ/. In Brazilian Portuguese, unstressed vowels are reduced to a lesser extent: final unaccented vowels are /i/, /u/, /a/ (realized as [ɐ]) and nonfinal unaccented vowels are /i/, /e/, /a/, /o/, /u/; furthermore, there is no phonemic accented /ə/.

Peculiar to Portuguese are the inflected infinitive and constraints on clitic placement. The inflected infinitive is the outcome of the Latin imperfect subjunctive. Since by regular sound change its forms became identical to the forms of the bare infinitive in the first- and third-person singular, the whole paradigm was reanalyzed as an infinitive with personal endings. Nowadays, it displays the following forms:

Singular	Plural
1. eu falar	falarmos
2. falares	falarem
3. ele/ela falar	falarem

(from the verb *falar*, 'to speak').

European Portuguese has retained a trace of the once pan-Romance Tobler – Mussafia Law, according to which no pronominal clitics could precede the verb when the latter was placed in sentence-initial position. Hence, we find in European Portuguese groups like *disse-mo*, 'he told me', to be compared with Spanish *me dijo*, or Italian *mi disse* (in Brazilian Portuguese this constraint does not hold; in general, clitics are preposed to a much larger extent than in European Portuguese; for example, they are preposed with infinitives: European Portuguese *buscar-me*, 'to look for me', Brazilian Portuguese *me buscar*).

The first attestations of Portuguese and Galician (which in ancient times were not yet distinct) date back to the twelfth century CE; Galician-Portuguese lyrics flourished especially at the court of King Denis (1279–1323).

Galician (Gallego)

Galician is spoken by about four million people, most of them living in the region of Galicia, in Spain, where it has the status of official language together with Spanish. It is intimately related to Portuguese, and the reasons why it is considered a separate language, rather than a vernacular variety, are ultimately political. In the archaic period, until the sixteenth century CE, there was only one literary variety, referred to as Galician-Portuguese; only in the course of the

sixteenth century did Galician separate from Portuguese, although some distinctions emerged, as a consequence of political developments, in the period from 1350 to 1500. The standardization of Galician was a slow and uncertain process, with conflicting tendencies either toward adherence to the Portuguese standard, or to a deeper distinction; only after the death of Franco, with the new Spanish constitution that recognized official status to minority languages, has Galician achieved full recognition.

Catalan (Catalá)

Catalan is spoken by about five million people, for the most part living in Catalunya, in the province of Valencia and in the Balearics, where Catalan is the official language together with Spanish; small Catalan-speaking communities are located in the South of France and in Northern Sardinia. Catalan is also the official language for about 31,000 speakers in the Republic of Andorra.

As already remarked, Catalan was in origin closer to Galloromance, and in particular to Occitan; also geographically, Catalan was closer to the area of the Provence: Catalan-speaking communities in the early Middle Ages were located in the South of France and in present-day Catalunya, and expanded to Valencia during the Reconquista; Mallorca and the Balearics were conquered by Catalan speakers in 1229. As a result of historical developments, Catalan nowadays builds a linguistic bridge between the Iberic and the Galloromance areas.

Features peculiar to Catalan are the vowel system, consisting of eight vowels— /i/, /e/, /ɛ/, /ə/, /a/, /ɔ/, /o/, and /u/— and the loss of diphthongization of Latin short /e/ and /o/ (but in there is evidence for early diphthongization in Medieval Catalan). Palatalization of initial /l/ is a feature of modern Catalan, having started in the sixteenth century (in e.g. *lleit*, 'milk', [ʎeit], cf. Spanish *leche*, French *lait*).

Catalan shares with standard Spanish the distinction between the auxiliaries *ser* and *estar*, although the extent to which *estar* is not as wide as in Spanish (*ser* can also occur when it means 'to be in a certain place', where Spanish has *estar*). An innovation of Catalan is the auxiliary use of *anar*, 'to go', with the infinitive, which yields a periphrastic preterit: *va cantar*, 'he sang', equivalent to Spanish *cantó*. The prepositional accusative did not, in origin, belong to Catalan; today, it has spread from Spanish to some extent, and has become the standard in the case when the object is a strong personal pronoun, or the indefinite *tots*, 'all', 'everybody', or in cases where word order could cause confusion between the subject and the object.

The first attestations of Catalan are somewhat later than those of Spanish and Portuguese, because for a long period Catalan poets used Provençal, which had a higher prestige. The first written document is a liturgical book dating to the beginning of the thirteenth century. In the course of the same century, the Catalan kings extended their reign to the Balearics, Valencia, and Murcia; consequently, Catalan acquired official status. Catalan was also the official language of the kingdom of Aragona, until unification with Castilla in 1479. Spanish then superseded Catalan as the official language; only in the nineteenth century did separatist movements give an important contribution to the maintenance and the standardization of the language.

References

Badia i Margarit, Antoni M. 1951. Gramática histórica catalana. Barcelona: Enciclopèdia Catalana.

Badia i Margarit, Antoni M. 1962. Gramática catalana, 2 vols. Madrid: Gredos.

Harris, Martin, and Nigel Vincent (eds.) 1988. The romance languages. London–Sydney: Croom Helm.

Meyer Lübke, W. 1890–1902. Grammatik der romanischen Sprachen. Leipzig: Fues.

Posner, Rebecca. 1996. The romance languages. Cambridge: Cambridge University Press.

Sala, M. 1982. El español de America. Bogotá: Instituto Caro Y Cuervo.

Sephiha, H.V. 1986. Le Judéo-espagnol. Paris: Entente.

Thomas, E.W. 1969. The syntax of spoken Brazilian Portuguese. Nashville: Vanderbilt University Press.

Williams, E.B. 1962. From Latin to Portuguese, 2 edition. Philadelphia: University of Pennsylvania Press.

Wright, R. 1982. Late Latin and early romance in Spain and Carolingian France. Liverpool: Francis Cairns.

Zamora Vicente, A. 1967. Dialectología española, 2nd edition. Madrid: Gredos.

SILVIA LURAGHI

Spectral Analysis

Speech production and speech acoustics are two main areas of study in experimental phonetics. Speech production is concerned with how the vocal organs move in relation to the linguistic message the speaker intends to deliver. Speech acoustics is concerned with the relationship between the moving vocal organs and the

acoustic signal that they create and also with the way in which the listener is able to recover the linguistic message from the acoustic signal. 'Spectral analysis' refers to any technique that transforms an acoustic signal into its 'frequency components'. Spectral analysis is an indispensable part of speech acoustics, and it has applications in areas such as clinical speech pathology and the development of systems for automatically synthesizing and recognizing speech.

Nature of an Acoustic Signal

An acoustic signal is created whenever an object moves, causing disturbances to the air pressure in its vicinity. These air pressure changes are propagated outward in all directions from the object and, if they are sufficiently large, a sound is heard when they reach a listener's ears. The loudness of the sound is related to the amount or 'amplitude' of air pressure change relative to the atmospheric pressure. Different speech sounds are heard because the moving vocal organs can be configured in a variety of ways and create different patterns of disturbance to the air pressure.

A 'speech pressure waveform' or simply 'waveform' is a record of the amplitude of air pressure changes as a function of time that result from speech production. A waveform provides much information about the differences between broad groups of speech sounds: in Figure 1, the pattern of air pressure variations is very different for the oral stop /t/ compared with the fricative /s/ or the vowel /æ/. But it is generally much less useful for showing the acoustic differences *within* any of these classes (i.e. stop, fricative, vowel), e.g. acoustic differences stemming from the different positions where the airstream is blocked or

constricted in the speaker's mouth (place of articulation). These phonetic distinctions generally only emerge acoustically by transforming a waveform into another kind of display, known as a 'spectrum', which displays the amplitude of air pressure variation as a function of frequency. It is this transformation that is at the core of spectral analysis.

What follows is the description of a type of signal known as a 'sinusoid', which is the building block of spectral analysis, and then a brief introduction to Fourier's theorem, which shows that *any* signal can be represented by the sum of sinusoids—thereby allowing the spectrum to be obtained for any waveform.

Sinusoids

Imagine a wheel with a number of scoops fixed to a water mill that is turned at a constant speed by a flowing river. Suppose a footpath runs alongside it so that half the wheel is below the footpath and the other half is above it (Figure 2), and that the radius of the wheel is one meter and exactly one revolution is completed every second. If at several points in time during the wheel's cycle, measurements are made of the height of one of the scoops above or below the path, the resulting plot is a sinusoid. Since (in this example) the wheel completes one revolution of the cycle every second, the 'frequency' of the sinusoid is said to be one cycle per second or 1 Hertz (abbreviated to 1 Hz). The sinusoid has an axis of amplitude (of height in the case) and an axis of time, but it is possible to represent the same information in a spectrum that has axes of amplitude and frequency. For this example, the spectrum of the sinusoid is a single vertical line, of amplitude 1 (meter), at a frequency of 1 Hz (Figure 3a)
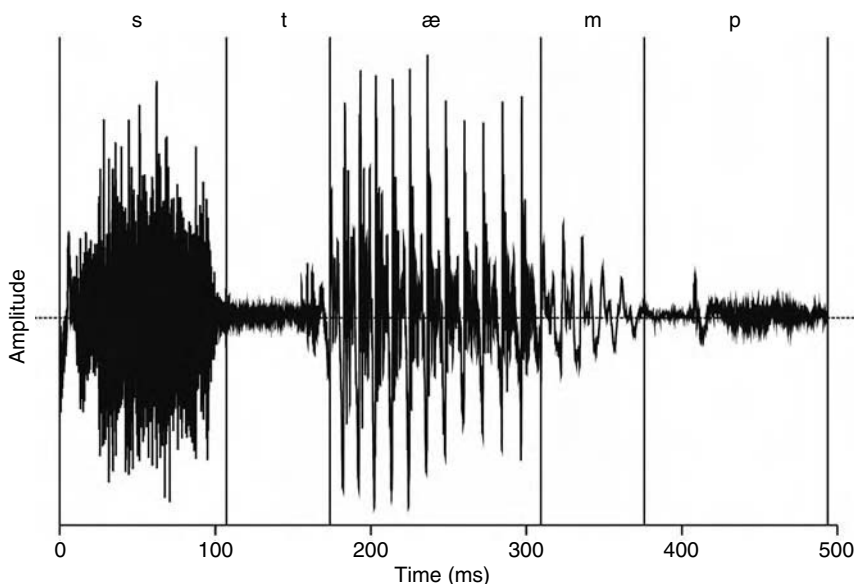
Figure 1. Waveform of the word 'stamp' produced by an adult male speaker of Australian English. The horizontal line corresponds to atmospheric pressure. The vertical lines show the approximate location of the boundaries between the speech sounds.
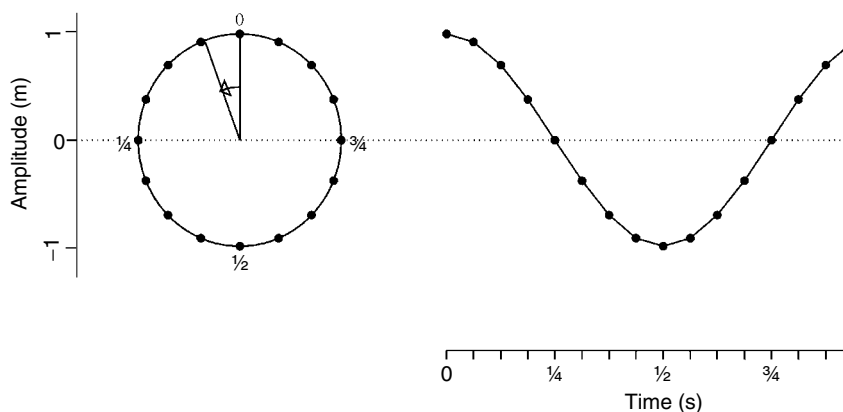
Figure 2. The circle represents a revolving wheel and the dots are one of the scoops at different time points on the wheel as it revolves. The horizontal dotted line is the path. The values around the circle are the times at which the scoop reaches that point in the wheel's cycle. A plot of the height of the scoop above or below the path as a function of time results in the sinusoid on the right.
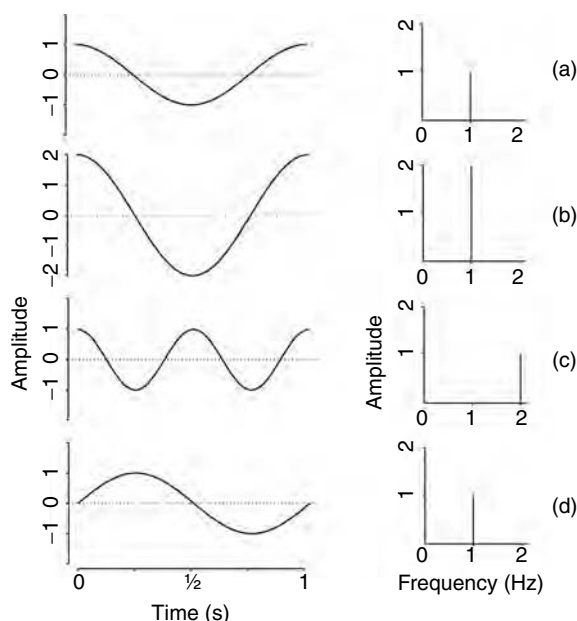
Figure 3. Waveform and spectrum of a sinusoid of amplitude 1 and frequency 1 Hz (a); amplitude 2 and frequency 1 Hz (b); amplitude 1 and frequency 2 Hz (c); and amplitude 1 and frequency 1 Hz and phase-shifted $\frac{1}{4}$ cycle earlier (d).

corresponding to the wheel of radius 1 unit that completes one revolution every second.

Now consider the effect on the resulting sinusoid of changing the configuration of the wheel, path, river, and scoop in three different ways. First, there might be a wheel of twice the size (in a different river) that happens to revolve at exactly the same speed as the one that has been considered so far. The resulting sinusoid and spectrum in Figure 3b show that the amplitude has doubled but the frequency is the same—therefore, the spectrum shows a vertical line of 2 units at a frequency of 1 Hz. Second, there might be yet another wheel that

has the same size as the one that was originally considered but that revolves at twice the speed. In this case, the amplitude is still 1 unit, but 2 cycles are completed every second—hence, the resulting frequency is 2 Hz (Figure 3c). The third type of change is one of 'phase', which refers to where in the cycle the measurements are first taken. For example, if the first measurement in Figure 2 is taken a quarter of a cycle earlier, when the scoop is level with the path and rising, the sinusoid is itself shifted a quarter of a cycle back in time (Figure 3d). Notice that the spectrum is unchanged, because neither the amplitude nor the frequency is affected by a change in phase. It turns out that as far as distinguishing acoustically between speech sounds is concerned, phase changes are more or less irrelevant; hence, they are considered no further in this discussion.

Periodic Waveforms and Fourier Analysis

The waveform at the top of Figure 4 is said to be 'periodic' because it has a basic pattern, or 'cycle', that repeats itself in time. The rate at which the pattern in a periodic waveform is repeated is known as its 'fundamental frequency'. This waveform repeats itself 2.5 times a second, and hence its fundamental frequency (which is abbreviated $F0$ or $f0$ and pronounced 'F-zero') is 2.5 Hz.

A listener hears a periodic waveform as having 'pitch'. The pitch is perceived to increase in approximately equal steps on the frequency scale up to about 1,000 Hz, and thereafter it increases approximately logarithmically. A note played on a musical instrument has a periodic waveform and an identifiable pitch, as do most vowels that are produced in speech. On the other hand, the sound produced from rustling leaves or the slamming of a door has an 'aperiodic' waveform with no discernible repeating pattern.

Accordingly, listeners do not hear them as having pitch in the same sense.

Fourier analysis, named after the French mathematician Joseph Fourier (1768–1830), is a technique that allows a waveform to be decomposed into a set of sinusoids. The meaning of 'decomposed' is that if the sinusoids that result from Fourier analysis are summed (i.e. played back at the same time), then the original waveform is exactly reconstructed. In Figure 4, the sum of the simple sinusoids *b*, *c*, *d* yields the complex waveform in *a*.

There is a second important property of Fourier analysis: the sinusoids that result from the decomposition of the type of periodic waveform shown in Figure 4a have frequencies at multiples of the periodic waveform's fundamental frequency. This is why the lowest frequency sinusoid that is produced from Fourier analysis in Figure 4, namely sinusoid *b* (known as the 'fundamental'), also has a frequency of 2.5 Hz, while the next two sinusoids higher in frequency (known as the '2nd and 3rd harmonics') have frequencies of 5

and 7.5 Hz for sinusoids *c* and *d*, respectively. The spectrum of a periodic waveform always consists of a set of discrete lines on the frequency axis and is known as a 'line spectrum'.

The examples so far have been of mathematically pure waveforms. As far as speech production is concerned, most vowels and many voiced consonants are produced with the vocal folds (vocal cords) opening and closing at a rapid rate (typically around 100 times per second or 100 Hz in adult male speakers and just below 200 Hz in adult female speakers), and this periodic vocal fold vibration results acoustically in a waveform that is sufficiently periodic that the harmonics are generally detectable in the spectrum. In Figure 5, there are peaks in the spectrum at intervals of about 90 Hz that are the result of the harmonics, giving rise to a sawtooth effect in the spectrum. This 90 Hz interval is predictable from the waveform in Figure 5 in the following way. The waveform shows that there are 4 cycles in 0.043 seconds. The fundamental frequency is given by the number of cycles per second, which in this case is 4/0.043 Hz or just over 90 Hz.

Voiceless sounds are produced without vocal fold vibration, and the airstream becomes turbulent as it is channeled at high speed through a narrow gap at the consonant's place of articulation. This turbulent airstream, which is audible as the hissing sound in the [s] of 'see' or the [ʃ] of 'she', results in an 'aperiodic' signal that has no cycle or repetitive pattern. In contrast

Figure 4. A periodic waveform and the three sinusoids that are the result of its Fourier analysis. *a*, *b*, *c*, *d* denote amplitude values on the waveforms at the same time point and their relationship is $a = b + c + d$.
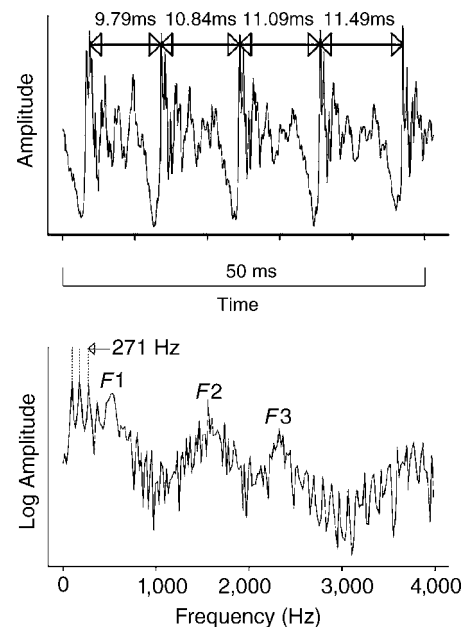
Figure 5. Top: Waveform of four pitch periods of the [æ] of 'stamp' from Figure 1 and their durations. Bottom: The spectrum of this waveform showing the first three harmonics (vertical dotted lines) and the frequency location of the first three formants (*F*1, *F*2, *F*3).

to a periodic signal, the spectrum of an aperiodic signal does not have equally spaced harmonic components but is 'continuous': that is, there are spectral components at *all* frequencies (Figure 6).

Formants

Consider the effect of tapping two glass jars, one long and quite wide and the other short and thin. These taps result in two quite different sounds (one sounds much lower in pitch than the other). The reason for these differences is that each jar has its own set of 'resonance frequencies' —these are the frequencies at which the air inside the jar is likely to vibrate at highest amplitude, once the air is set in motion by the tap against its side. The frequencies at which the jar resonates with highest amplitude depend predominantly on the jar's shape. So too in speech production. If you set yourself to produce an [æ] vowel (analogous to one of the jars) and flick a finger against your throat, you will hear a muted sound that decays quickly. If you do it again but set yourself to produce an [i] vowel, you will hear a quite different sound. Since [æ] and [i] are produced with very differently shaped vocal tracts, the set of resonance frequencies, which in the case of the acoustic analysis of voiced sounds are known as 'formant frequencies,' are also quite different.

In the spectrum of a speech sound, the formant or resonance frequencies show up as amplitude peaks and are labeled $F1$ (the first formant frequency), $F2$, $F3$, …, Fn (Figures 5 and 6). The first two formant frequencies provide listeners with the most important information for distinguishing between phonetically different vowels. They are proportional, respectively, to how widely open the lips are and the position between the lips and vocal folds of greatest narrowing in the vocal tract.
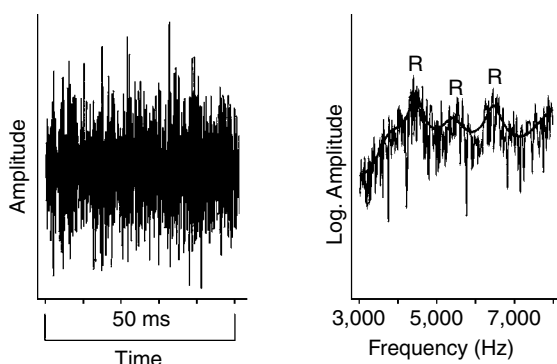
Figure 6. Waveform of the section of [s] from 'stamp' in Figure 1 and its spectrum (right). A trend line is superimposed on the spectrum. Three of the resonance frequencies for [s] between 3,000 and 7,000 Hz are marked as R above the spectrum.

Source–Filter Theory of Speech Production

Speech production is often characterized as the composition of a 'source' and a 'filter'. The source causes the air in the vocal tract to vibrate: for voiced sounds, the source is the vibrating vocal folds, and for voiceless sounds it is a turbulent airstream. In speech production, the shape of the vocal tract acts as a filter because it modifies the effects of the source. The different contributions to the spectrum of the source and filter can be seen in Figures 5 and 6: the source results in short-term variation that rides up and down on a longer-term trend line (whose peaks are the formants), which is the acoustic effect of the filter.

In speech production, the source and filter are largely independent of each other because they arise from processes that can themselves be independently varied. In voiced speech, for example, a speaker can produce the same vowel but can vary the pitch (as when singing an [æ] vowel on a rising pitch). The acoustic consequence of this independence of the source and filter is that the formant frequencies are negligibly affected by changes in the harmonic structure. If the same [æ] vowel as in Figure 5 were produced at twice the pitch, the frequency interval between the harmonics (short-term variation) would double, but the shape of the long-term trend line due to the filter (vocal tract shape) would be largely unaffected.

Spectrograms

A spectrogram is a three-dimensional representation that shows how spectral information changes in time. It is calculated by dividing the waveform into equal time sections or 'windows' and applying a Fourier analysis to the waveform within each window. Each analysis produces a spectrum of exactly the kind that has been considered so far but in which the amplitude is represented by a gray scale (white is zero amplitude; black is maximum amplitude). In Figure 7, which is a spectrogram of the waveform in Figure 1, the first three formants in the vowel are clearly visible at about 700, 1,600, and 2,300 Hz.

Figure 7 is an example of what is often known as a 'wideband spectrogram', and it has a 'bandwidth' of 300 Hz. Any two components whose frequency difference is less than the bandwidth are indistinguishable on the frequency axis. Since for this speaker the fundamental frequency for the [æ] vowel was estimated at 90 Hz (Figure 5), the separate harmonics, which occur at 90 Hz intervals, are indistinguishable on this wideband spectrogram, because their frequency separation is less than the 300 Hz bandwidth. The 'temporal resolution' of a spectrogram, which can be used to determine whether two events are separately visible on the time axis, can be estimated from the reciprocal of the band-
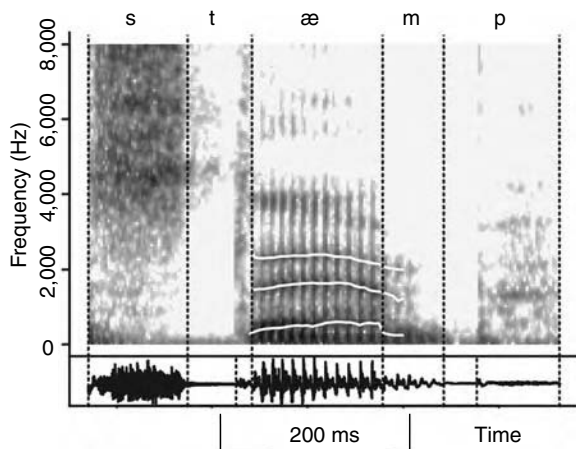
Figure 7. Wideband spectrogram and waveform of 'stamp' in Figure 1. The vertical dotted lines mark the approximate location of boundaries between the speech sounds. The white lines on the spectrogram are the first three formant frequencies.

width. For this wideband spectrogram, the temporal resolution is 1/300 second or just over 3 milliseconds. Accordingly, the individual vocal fold closures for the [æ] vowel are separately visible as vertical lines or 'striations' (Figure 7), since the interval between them, which is approximately 1/90 second or 11 milliseconds, is greater than the spectrogram's temporal resolution.

Figure 8 is a 'narrowband spectrogram' that was calculated with a 45 Hz bandwidth. In this case, the vertical striations for [æ] are no longer visible in the vowel, because the temporal resolution, which is given by 1/45 second (just over 22 milliseconds) is greater than the interval between vocal fold closures. But the harmonics *are* visible because the frequency difference between them (90 Hz) is greater than the 45 Hz bandwidth. This type of spectrogram has often been used for measuring the extent and direction of changes in pitch. From Figure 8, it is immediately clear that the word was produced on a falling pitch.

Cepstrally and LPC Smoothed Spectra

Speech scientists often need to be able to estimate the separate contributions to a speech signal of the source and the filter. Two techniques for doing this, which can only be applied to signals that have been digitized on a computer, are 'cepstral analysis' and 'linear predictive coding' (LPC).

The starting point for cepstral analysis is the observation that was made earlier: the spectrum of a speech sound is made up of a rapidly changing part that is due to the source and a slowly varying trend line that is the result of the filter. Consider in the light of this the problem of separating two sinusoids that are merged together in a waveform, one of which varies rapidly and the other slowly (at frequencies of 100 and 5 Hz,
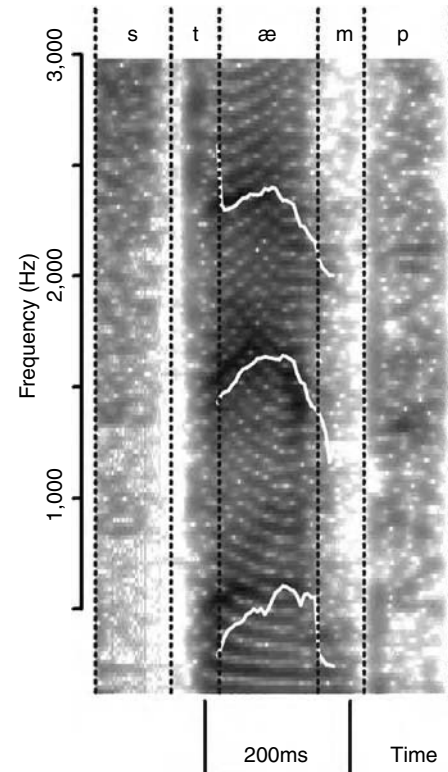
Figure 8. Narrowband spectrogram of 'stamp' in Figure 1. The white lines are the first three formant frequencies.

for example). This separation could be achieved by applying Fourier analysis, resulting in a spectrum with two separate lines at these frequencies. Therefore, if Fourier analysis were applied to a spectrum of speech, then the result, known as a cepstrum, should be a separation of the rapidly and slowly varying components, i.e. of the source from the filter. Since they appear in different parts of the cepstrum, the source can be quite easily removed, leaving only the contribution from the filter.

If Fourier analysis is applied once more, but this time to the cepstrum with the source removed, the result is another spectrum: it very much resembles the original spectrum of the speech signal, but it is considerably less bumpy, precisely because the contribution from the source component has been mostly removed. One of the main advantages of this spectrum, which is known as a 'cepstrally smoothed spectrum', is that it is much easier to see the formant frequencies, since they are not confounded with the short-term effects of the source (Figure 9). The part of the cepstrum that includes information about the filter, known as 'cepstral coefficients', provides considerable information about the acoustic differences between phonetically distinct sounds. For this reason, cepstral coefficients are often used in systems for automatic speech recognition.
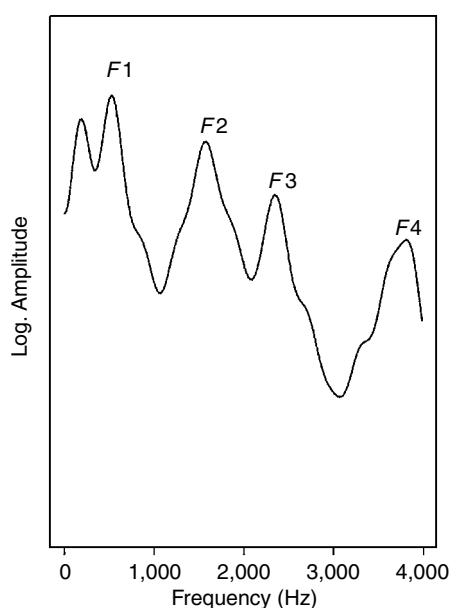
Figure 9. Cepstrally smoothed spectrum of the waveform in Figure 5 showing the frequencies of the first four formants.

The basis for the LPC analysis of speech, which in other fields is sometimes known as 'autoregressive analysis', is as follows. If a speech signal is the result of passing a source (vibrating vocal folds or a turbulent airstream) through a filter (representing the shape of the vocal tract), then it can be shown that the same signal is equal to the sum of several 'scaled' and 'delayed' versions of itself added to the source. 'Scaled' means that the speech signal's amplitude is linearly changed: the amplitude of the entire signal might be halved, for example. 'Delayed' means that the entire speech signal is shifted back in time: if the signal is delayed by a fraction of time, then all its values start a fraction of a time earlier. The factor by which each delayed signal is scaled is dependent on the characteristics of the filter and therefore on the shape of the vocal tract. Hence, a model of the vocal tract shape (filter) that could have

given rise to an acoustic speech signal can be reconstructed by estimating the different scale factors on successively delayed versions of that signal. This is what LPC tries to do. The scale factors are known as 'LPC coefficients'.

The Fourier analysis of the LPC coefficients results in an 'LPC-smoothed spectrum'. As with cepstral analysis, the spectrum is smooth because the contribution from the source is minimal. LPC analysis also allows the formant frequencies to be estimated, because there is a direct correspondence between them and the LPC coefficients. The formant frequencies in the spectrograms in Figures 7 and 8 were automatically tracked using LPC analysis.

References

Fant, G. 1960. The acoustic theory of speech production. The Hague: Mouton.
———. 1973. Speech sounds and features. Cambridge, MA: MIT Press.
Flanagan, J. 1972. Speech synthesis, analysis, and perception. New York: Springer-Verlag.
Harrington, J., and S. Cassidy. 1999. Techniques in speech acoustics. Dordrecht: Kluwer.
Johnson, K. 1997. Acoustic and auditory phonetics. Cambridge, MA: Basil Blackwell.
Kent, R., and C. Reid. 1992. The acoustic analysis of speech. San Diego: Singular.
Ladefoged, P. 1962. Elements of acoustic phonetics. Edinburgh: Oliver & Boyd.
Lehiste, I. 1967. Readings in acoustic phonetics. Cambridge, MA: MIT Press.
Lindblom, B., and S. Sundberg. 1971. Acoustical consequences of lip, tongue, jaw and larynx movement. Journal of the Acoustical Society of America 42. 830–43.
Olive, J., A. Greenwood, and J. Coleman. 1993. Acoustics of American English speech: a dynamic approach. New York: Springer-Verlag.
Pickett, J. 1998. The acoustics of speech communication. Boston: Allyn & Bacon.
Stevens, K. 1998. Acoustic phonetics. Cambridge, MA: MIT Press.

JONATHAN HARRINGTON

Speech Acts

In distinguishing between constative and performative language, the twentieth-century English philosopher J.L. Austin initiated 'speech act theory', which has influenced linguistics and literary criticism. Constative language refers to things; performative language does things—it is speech in action.

Constative language reports on reality. We evaluate constative language as true or false. 'The snow is white' is an example, as is the following: 'All men are mortal; Socrates is a man, so Socrates is mortal'. The first example describes a visual phenomenon. The second details a logical relation's

reality and describes Socrates as a man and all men as mortal.

Both examples are statements with a determinable sense and reference, usually called 'meaning'. In addition to simply conveying their meaning, statements may be uttered to bring about effects in listeners or readers. The statement 'There is a hungry lion in the room' produces fear in the room's human occupants, in addition to informing them about the lion. Although clearly a constative sentence, this sentence accomplishes a nonlinguistic speech act: it frightens.

Sentences may be uttered simply to say meaningful words, but in saying these words, a certain effect may be achieved. Both cases are speech acts, but whereas the first aims to achieve a linguistic effect—namely, that the listener understands what is said—the second may aim at nonlinguistic effects—for example, evoking a certain emotional response. Austin refers to constative sentences as 'locutions' and to utterances intended to achieve a different effect as 'illocutions', and the particular intent of illocutions is labeled 'illocutionary force'. A sentence that mentions a person allows us to understand to which person the sentence refers. If, in addition, the sentence commands that we despise or admire that person, the sentence is an illocution. An illocution that not only aims to achieve an effect but is also successful at achieving it is called a 'perlocution'. All commands are thus illocutions, but only commands resulting in obedience are perlocutions.

Illocutionary force is distinct from locutionary sense, reference, and meaning, even though illocutions necessarily involve locutions. Consider the following illocution: 'Impeach George W. Bush, President of the United States: he lied about Iraq's weapons of mass destruction!' This example has illocutionary force whether the words 'he lied about Iraq's weapons of mass destruction' are or are not legally meaningful (is Bush's lying an impeachable offense?) or do or do not have a referent (did Bush lie?). If the example's illocutionary force results in Bush's impeachment, it has achieved a perlocutionary act.

What Austin calls 'performatives' constitute a particular class of illocutions. The typical performative includes a singular first-person pronoun as the subject ('I') and an active verb in the present tense indicative. 'I promise to pay' and 'I declare the meeting over' are examples. 'I declared the meeting over', a constative locution, reports, but does not perform, a past act and can be judged true or false.

Performatives are 'felicitous' or 'infelicitous' rather than true or false. Felicitous performatives conform to particular conditions, which are often institutional conventions, whereas infelicitous performatives do not. A classic example comes from the marriage ceremony. 'I pronounce you husband and wife' does not describe two people as being husband and wife: it makes them husband and wife, but only in the context of certain felicity conditions. If the person uttering the performative lacks authority to perform marriages, the man and the woman are not made husband and wife.

Austin lists general felicity conditions for performatives: that recognized conventions specify who should pronounce the performative, by what procedures, and to what effect; that the participants enact these conventions correctly and completely; and that the person uttering the performative intends it seriously. If, after the jury has returned a 'not guilty' verdict, the judge says to the accused, 'I sentence you to death,' we would have an infelicitous performative because of conventions being violated. If John Doe says, 'I promise to pay' without intending to keep the promise, the performative is infelicitous. Austin does not consider promises or death sentences uttered by an actor playing a character on stage to be performatives because they are not intended seriously: they are fictions.

Speech acts' relations to intention are matters of ongoing debate. In *Limited Inc* (1988), the philosopher Jacques Derrida claims that Austin privileges intention as the anchoring center of the context that grants a performative felicity. Linguistic signs, i.e. words, sentences, and so on, must be able to function in their speakers' or writers' absence. Listeners can understand a voice echoing in a canyon even if the speaker died during enunciation. The voice repeats its message despite the speaker's absence. Signs can be iterated, or repeated, beyond the context of their production. This inherent repeatability, argues Derrida, guarantees that speech acts cannot be fully explained by any given context and cannot wholly depend on intention. A speech act always has a context, but iteration prevents that context from ever completely determining the speech act's meaning. As written or verbal signs, performatives function in the radical absence of their writers or speakers and, thus, their intentions.

Michelle Rosaldo's essay in *Language in society* (1982), 'The things we do with words: Ilongot speech acts and speech act theory in philosophy', argues that a focus on intention, especially in relation to promises, biases speech act theory toward modern individualist ideology by emphasizing autonomous psychological states. Miller's *Speech acts in literature* (2001) extends Derrida's arguments to claim that Austin's lectures at once display and subvert sexist and imperialist ideologies. Miller details how Austin's examples tie intention to a white male speaker's sovereignty. However, Miller argues, Austin unintentionally dismantles the constative/performative opposition, opens the possibility of performatives preceding and establishing their own felicity conditions, and discloses

other features of speech acts that militate against the very ideologies that Austin's arguments harbor.

Countering arguments such as Derrida's, the philosopher John Searle maintains that intentionality is a legitimate foundation of speech act theory. Searle differentiates between intentionality as a prelinguistic feature of human and animal life and intentionality in language, which is ultimately derivative of intentionality as a biological endowment. Language and speech acts are to be explained in terms of intentionality, rather than intentionality being explained in terms of language. Reasserting context as a controlling factor, Searle extends Austin to define types of illocutions (to request, to congratulate, and so on) and their conditions of felicity.

References

Austin, J.L. 1961. Philosophical papers, Oxford and New York: Oxford University Press, 2nd edition, 1970, 3rd edition, 1979.

———. 1980. How to do things with words: the William James lectures delivered at Harvard University in 1955. Oxford and New York: Oxford University Press.

Butler, Judith. 1997. Excitable speech: a politics of the performative. New York and London: Routledge.

Derrida, Jacques. 1988. Limited Inc. Evanston: Northwestern University Press.

Miller, J. Hillis. 2001. Speech acts in literature. Stanford: Stanford University Press.

Petrey, Sandy. 1990. Speech acts and literary theory. New York and London: Routledge.

Rosaldo, Michelle Z. 1982. The things we do with words: Ilongot speech acts and speech act theory in philosophy. Language in Society 11(2). 203–37.

Searle, John R. 1969. Speech acts: an essay in the philosophy of language. Cambridge and New York: Cambridge University Press.

———. 1977. Reiterating the differences: a reply to Derrida. Glyph 1. 198–208.

———. 1979. Expression and meaning: studies in the theory of speech acts. Cambridge and New York: Cambridge University Press.

———. 1983. Intentionality: an essay in the philosophy of mind. Cambridge and New York: Cambridge University Press.

Tsohatzidis, Savas L. (ed.) 1994. Foundations of speech act theory: philosophical and linguistic. London and New York: Routledge.

ROBERT S. OVENTILE

See also **Austin, John Langshaw; Ideology and Language; Searle, John**

Speech Perception

The speech stream is a continuously varying signal that, contrary to the listener's impression of silence between words, does not contain any pauses. The primary goal of research in speech perception is to illuminate the way in which the listener converts this continuous signal into a sequence of discrete, meaningful units. This process draws on perceptual, linguistic, and cognitive factors.

At least three major issues need to be taken into account when considering the mapping between acoustic parameters and linguistic units. The first concerns the segmentation of the speech stream. Speech is a highly efficient means of communication in which multiple layers of information are transmitted in parallel. Compare, for example, the words 'sea' and 'Sue'. The former consists of an [s] consonant followed by a vowel [i], which is pronounced with unrounded lips, whereas the latter consists of the same initial consonant followed by a rounded vowel [u]. During production of the initial consonant in 'Sue', lip rounding will start in anticipation of the upcoming rounded vowel, a process known as anticipatory coarticulation. As a result, the acoustic properties of [s] preceding [u] will be different from those of [s] preceding [i]. Thus, the pronunciation of the initial sound in the above words provides acoustic cues about the initial sound itself and about the immediately following sound at the same time. This parallel transmission of information can lead to a segmentation problem in that it is difficult or impossible to chop up the speech signal in chunks that correspond to only a single speech sound. In other words, it is difficult—if not impossible—to tell exactly where the consonant ends and the vowel begins.

A second related issue concerns the lack of linearity. Although the listener perceives the speech signal as a linear sequence of units (the size of which is still under debate), the acoustic cues to these units do not necessarily occur in a corresponding left-to-right order in the speech stream. Consequently, acoustic properties appearing later in the speech stream may carry information that is crucial for the identification of an earlier occurring speech sound.

The third issue concerns variability or the lack of invariance. The properties of the acoustic signal that are thought to elicit perception of speech sounds (the acoustic 'cues') are never exactly the same. Sources of variability are, for example, differences in vocal tract size, speaking rate, phonetic context, emphasis, and intonation; these can significantly affect acoustic parameters of the speech signal. For example, the most salient frequencies ('formants') of the vowel in the word *heed* are approximately 300 and 2,300 Hz for an adult male and 500 and 3,100 Hz for a child. However, despite these very different formant frequencies caused by differences in vocal tract size, listeners perceive both utterances as containing the vowel [i]. This illustrates what is known as the invariance problem: acoustic cues to a particular speech sound may not be constant but may instead vary according to the circumstances under which they occur. Thus, listeners have to compensate for such differences, or 'normalize the input'. One of the primary issues in speech perception, then, is how the listener achieves an invariant percept despite great variability in the acoustic input.

Speech perception experiments usually involve manipulated natural speech or synthetic computer-generated speech. Systematic manipulation of individual attributes of speech enables researchers to determine which acoustic properties are necessary and sufficient cues for a particular percept. In applying this methodology to consonants, early research in the 1950s revealed that English listeners use two primary acoustic cues for determining where in the mouth exactly a consonant is articulated: the frequency of the release burst and the formant transitions from the consonant into the following vowel. When the researchers exposed listeners to consonant–vowel sequences, keeping the burst frequency constant, the listeners perceived a [p] when an [i] was following, but a [k] when an [ɑ] was following. Thus, the exact same cue can participate in different perceptions. In highlighting that individual acoustic attributes are highly context dependent, these experiments also exposed the invariance problem.

Debate continues concerning the extent to which the perception of speech involves the use of biological mechanisms evolved especially for speech. Evidence that speech sounds are perceived differently from their nonspeech analogs was first presented in the early 1960s. These studies specifically examined the way in which these two types of sounds are identified and discriminated. Most types of stimuli (e.g. musical tones, colors) are much better discriminated than they are identified. The greater the physical difference between two stimuli, the better their discrimination. This was shown for nonspeech sounds as well. However, this was not true of certain speech sounds, most notably

stop consonants such as [b] or [d]. Discrimination of these sounds was not any better than their identification. For example, in a typical experiment with synthetic speech, an important formant frequency of the consonant was manipulated. Typically, [b] has a formant at 1100 Hz, whereas the equivalent formant of [d] lies at 1,800 Hz. With the help of a speech synthesizer, researchers were able to produce sounds with formants that lie somewhere in between. When listeners were asked to identify such intermediate consonants, it turned out that listeners seem to use a particular frequency as the break-off point: all consonants whose formant exceeded this threshold frequency were identified as [d], and all others were identified as [b]. In another experiment, listeners were presented with a pair of consonant stimuli and asked to tell whether the stimuli were same or different. When the formants of consonants both exceeded or both undercut the threshold frequency, the listeners seemed to answer at random. If one consonant fell into the [d] range and the other into the [b] range, however, listeners were consistently able to categorize them properly, even if the formant frequencies were close together. This pattern of results is known as categorical perception.

It was originally thought that categorical perception occurred only with speech sounds and not with nonspeech sounds, and this would suggest that the perception of speech engaged specialized mechanisms. However, later experiments with carefully controlled nonspeech materials have shown patterns of categorical perception as well. In addition, animals including the chinchilla, macaque, and Japanese quail have also been shown to have human-like categorical perception. These data concerning categorical perception do not readily support postulation of perceptual mechanisms that were specially evolved or adapted for speech. Instead, they seem to favor an interpretation based on general auditory mechanisms and psychoacoustic sensitivity.

The finding of categorical perception suggested that listeners did not perceive any differences between stimuli belonging to the same category. This, however, may have been based on the particular response categories that were typically used in identification and discrimination experiments (e.g. /b/ or /d/, 'same' or 'different', respectively). The use of more sensitive response measures reveals that even though listeners do assign the same label or fail to distinguish between stimuli of the same category, they are in fact aware of subtle differences. These findings indicate that although categories play an important role in the perception of speech, they are not monolithic but have an internal structure to which listeners are sensitive.

The importance of speech categories leads to the question of how they are established. This debate

centers around the issue of whether speech categories are innate or result from exposure to the ambient language. Crucial evidence in this debate is typically drawn from perception experiments with infants. Findings from discrimination experiments with infants as young as one month old suggest that they divide a speech continuum in a way very similar to adults, with two clearly defined categories and a sharp boundary at the adult location. Additional research has shown that infants up to approximately six months of age can not only discriminate speech categories from their native language but also from just about any other language, as well. However, in the second half of the first year of life, infants seem to lose their sensitivity to nonnative distinctions. Presumably, the decrease in sensitivity to contrasts that do not play a role in the native language allows for an increase in attention to other aspects of the speech signal that play a role in word learning, such as sentence structure and intonation. Acquisition of speech categories can thus be understood as the result of the interaction between initial psychoacoustically based sensitivities and an increasing awareness of the structure of the language to be learned.

Because the perception of speech draws on many sources of knowledge, theories of speech perception often account for only a few of its components. Three general classes of models may be distinguished. The motor theory of speech perception deals with acoustic variability by claiming that the listener has specialized neural mechanisms to convert the speech signal into invariant representations of articulatory gestures. These articulatory gestures are the object of speech perception; i.e. this theory assumes that the listener attempts to faithfully reconstruct how the perceived speech sounds were articulated by the speaker. The theory of acoustic invariance claims that invariant acoustic properties do reside in the speech signal. By using specialized neural mechanisms, the listener directly extracts these invariants from the speech signal and maps them onto phonetic features. Finally, pattern recognition models claim that speech perception is much like statistical pattern classification. No specialized mechanisms are required. Instead, the unit of recognition and the structure of categories (e.g. based on prototypes or exemplars) are determined by the nature of the speech signal and general properties of the mammalian auditory system.

There is a growing recognition that a detailed analysis of the speech signal alone will not be sufficient to obtain a genuine understanding of the way in which speech is perceived. Consequently, researchers have started to incorporate findings from additional areas. One such area concerns the way in which the speech signal is transformed by the auditory system. Although the measurements used in phonetic analysis typically represent frequency along a linear scale, it is known that the auditory system warps the signal such that its ultimate representation is more nearly logarithmic in nature. A thorough understanding of these transformations at the auditory periphery and higher levels along the auditory pathway may well have significant implications for the current view of acoustic cues and their variability.

A second area concerns the relation between the speech signal and higher levels of organization of the grammar. The first few decades of research on speech perception emphasized the cataloguing of acoustic cues, that is, the acoustic information that listeners need to extract from the speech signal to recognize individual speech sounds. Since the 1980s, however, increased interest in the way words are recognized has led to research on what is often referred to as spoken or auditory word recognition. A central issue for this research area is if and to what extent 'higher-level' linguistic and cognitive information that is not present in the speech signal contributes to word recognition. Research has shown that when presented with a sequence of speech sounds containing an ambiguous initial consonant, listeners will classify that consonant such that the entire string will result in an existing word instead of a nonword. For example, listeners will report hearing 'beef' rather than 'peef' when presented with the string 'eef' preceded by a sound that is ambiguous between /b/ and /p/. Conversely, listeners will classify the same ambiguous initial consonant as /p/ rather than /b/ when it is followed by 'eace'. Findings such as these are often considered as evidence that lexical information (knowledge about what constitutes a word) affects the listener's interpretation of acoustic–phonetic information. The extent to which there is feedback from higher levels of linguistic representation, such as the lexicon, is still very much under debate.

References

Gernsbacher, Morton Ann (ed.) 1994. Handbook of psycholinguistics. San Diego, CA: Academic Press.

Goodman, Judith C., and Howard C. Nusbaum (eds.) 1994. The development of speech perception: the transition from speech sounds to spoken words. Cambridge, MA: MIT Press.

Hardcastle, William, and John Laver (eds.) 1997. The handbook of phonetic sciences. Oxford and Cambridge, MA: Blackwell.

Harnad, Steven (ed.) 1987. Categorical perception: the groundwork of cognition. Cambridge: Cambridge University Press.

Jusczyk, Peter. 1997. The discovery of spoken language. Cambridge, MA: MIT Press.

Kuhl, Patricia. 1989. On babies, birds, modules, and mechanisms: a comparative approach to the acquisition of vocal communication. The comparative psychology of audition: perceiving complex sounds, ed. by Robert J. Dooling and Stewart H. Hulse. Hillsdale, NJ: Lawrence Erlbaum Associates.

Liberman, Alvin M. 1996. Speech: a special code. Cambridge, MA: MIT Press.

Marslen-Wilson, William (ed.) 1989. Lexical representation and process. Cambridge, MA: MIT Press.

Miller, Joanne L., Raymond D. Kent, and Bishnu S. Atal (eds.) 1991. Papers in speech communication: speech perception. Woodbury, NY: Acoustical Society of America.

Moore, Brian C.J. (ed.) 1995. Hearing. San Diego, CA: Academic Press.

Nygaard, Lynne C., and David B. Pisoni. 1995. Speech perception: new directions in research and theory. Speech, language, and communication, ed. by Joanne L. Miller and Peter D. Eimas. San Diego, CA: Academic Press.

Schouten, Martin E.H. (ed.) 1987. The psychophysics of speech perception. Dordrecht: Kluwer.

Warren, Richard M. 1999. Auditory perception: a new analysis and synthesis. Cambridge: Cambridge University Press.

ALLARD JONGMAN

Speech Processing

Speech processing (SP) is used to facilitate the communication between humans and computers. A system that can converse with a human in a natural and unrestricted way has long been seen as the quintessential challenge in computing.

As a field, SP encompasses a wide variety of disciplines and technologies. From a scientific perspective, it draws on research from Signal Processing, Acoustics, Linguistics (Phonology, Syntax, Pragmatics, etc.), and other areas.

To present the main concepts of SP, we will be introducing a hypothetical computer information system, Sprach (the reader may think of an automated air travel reservation agent or a museum information system). The architecture of Sprach is shown in Figure 1.

We will describe how a user can engage in a conversation with the information system. The sample transcript of the dialogue is shown in Figure 2.

Speech Recognition

Jill, the user of the computer system, can communicate with Sprach using either a microphone or a telephone interface. In both cases, a speech recognizer is needed to convert the acoustic signal into a sequence of words. Given that a telephone samples its input signal at a lower frequency, the quality of speech recognition is lower than with a microphone.

Speech recognizers are typically designed to solve two distinct problems: isolated word recognition and continuous speech recognition. Recognizing continuous speech is more difficult than isolated speech recognition due to the presence of disfluencies (e.g. false starts or interjections such as 'uh' or 'hmm') and prosody effects (e.g. intonation and rhythm). The accuracy of speech recognition also depends on the size of the vocabulary used. It is much easier to tell
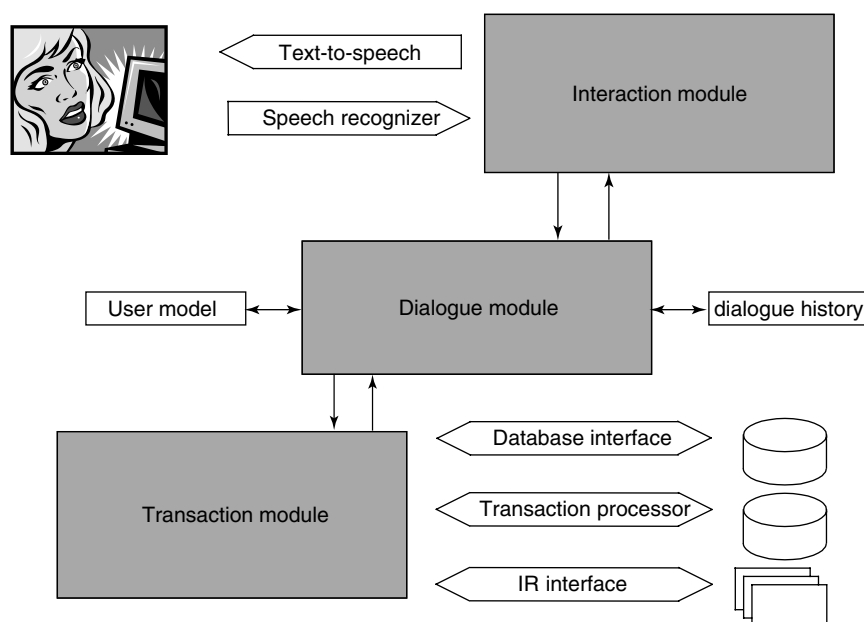
Figure 1. Sprach: a generic speech-based information system.

Jill: I would like to fly to Seattle on Friday.
Sprach: I am looking for flights from Detroit to Seattle on Friday,
          March 2. What time of the day do you want to leave?
Jill: I would rather fly early in the morning.
Sprach: American flight 143 leaves Detroit at 7:05 AM and arrives in
          Seattle at 9:28 AM. A ticket costs $590.00. Should I book
          one first-class ticket using your Discover card?
Jill: Yes, please.

Figure 2. A sample dialogue between Jill and Sprach.

apart individual digits from 0 to 9 than to recognize the names of all employees in a given company. Given the variability in speech characteristics among speakers, most practical speech recognizers also need to be trained for a particular user.

One mathematical formalism underlying speech recognition is the noisy channel model. It is used to guess the most likely utterance, given an acoustic signal. The most likely utterance $\hat{W}$ among the space Λ of all utterances $W \in \Lambda$ is identified according to the formula:

$$\hat{W} = \arg \max_{W \in \Lambda} P(W|A) = \arg \max_{W \in \Lambda} P(A|W)P(W)$$

In the formula, $P(A|W)$ is called the acoustic model and is used to estimate the most likely mappings between the acoustic signal A and the input utterance. The term $P(W)$ is known as the language model and is used as a measure of the grammaticality of a given candidate utterance W. Other techniques used in speech processing are Hidden Markov Models, Dynamic Time Warping, and Gaussian Mixture Models. For additional reading, consult the book by Huang et al.

Dialogue Management

Sprach's dialogue module shown in Figure 1 requires two locations to store the history of the interaction with the user. One of them is long-term memory (or user model) in which user preferences and information extracted from past dialogues is stored. For example, long-term memory may contain the city where Jill lives or her favorite airline. On the other hand, the short-term memory is used to represent choices that can be derived from the current dialogue, e.g. the day and time of the desired flight.

In practical speech-based dialogue systems, dialogue interactions between users and the system are often preprogrammed. Recently, VoiceXML has emerged as a standard for encoding information for such applications. VoiceXML is used to annotate computer–human dialogues. Some sample tags are: <MENU> : a list of options, <FORM> : a mechanism for user input, and <EMP> : for emphasized speech. VoiceXML and equivalent markup language facilitate the development of practical speech-based systems.

Back End

The back end of the system is where transactions are processed. Our sample system contains three back ends — an information retrieval (IR) interface, a database system interface, and an interface to a transaction processing component. The IR interface is used to retrieve documents (e.g. e-mail messages) from a server. The transaction processing and database interfaces can be used for a large variety of applications such as hotel or air travel reservation systems.

Text-to-Speech Generation

When Sprach has decided *what* to say, it still has to figure out *how* exactly to utter it. A text-to-speech (TTS) component is used for that purpose. The input to TTS is most often a grammatical sentence, either retrieved from a sentence dictionary or produced by text generation from a logical form. The input sentence is decomposed graphemes (textual units) which are then converted into phonemes (phonetic units). The process of grapheme-to-phoneme translation is often ambiguous, especially in languages with inconsistent pronunciation such as English. In English, the same phoneme /f/ can be generated from multiple graphemes (f, ph, gh, and ff), while the same grapheme can be translated to more than one phoneme (e.g. gh can be /g/, /f/, or silent, among other pronunciations). The two main approaches used in grapheme-to-phoneme translation are rule-based (using linguistically tested translation pairs) and corpus-based (derived from large corpora).

To make the output sound natural, TTS systems make use of intonational (prosody) information. Prosody features are expressed by controlling input parameters of the speech synthesizer. Some prosody features such as pauses can be inferred within a certain margin of error from the syntactic structure of the sentence (e.g. the presence of semicolons, commas, and other punctuation). Features such as timber, dialect, and speech rate are also often included in the generation process, mostly to produce voices that seem to be of different characters.

Evaluation

The speech recognition research community has adopted a set of evaluation metrics that are used to gauge progress over the years. One of these metrics is the error rate, which is simply the percentage of word insertions, deletions, or substitutions that are needed to convert the output of the speech recognizer to the correct text. State-of-the-art recognition systems achieve around 10% word error rates on unconstrained spontaneous speech with a vocabulary of more than 10,000 words. In other words, in a given sentence, one in ten words is incorrectly identified. However, when the text

is constrained, word error rate declines dramatically to approximately 4% for continuous speech with a 1,000-word vocabulary and less than 1% for isolated word recognition.

Evaluating speech generation is significantly more difficult and remains a research challenge.

References

Cole, Ronald, Joseph Mariani, Hans Uszkoreit, Annie Zaenen, and Victor Zue (eds.) 1997. Survey of the state of the art in human language technology. Cambridge: Cambridge University Press.

Huang, Xuedong, Alex Acero, and Hsiao-Wuen Hon. Spoken Language Processing: A Guide to Theory, Algorithm and System Development. Upper Saddle River, NJ: Prentice-Hall.

Jelinek, Frederick. 1998. Statistical methods for speech recognition. Cambridge, MA: MIT Press.

Jurafsky, Daniel, and James Martin. 2000. Speech and language processing. Englewood Cliffs, NJ: Prentice-Hall.

Pallett, David S., Jonathan G. Fiscus, John S. Garofolo, Alvin Martin, and Mark A. Przybocki. 1998 Broadcast News Benchmark Test Results. Proceedings of the DARPA Broadcast News Workshop, 1999.

www.voiceml.com. VoiceXML markup language (2001).

DRAGOMIR R. RADEV

Speech Processing (Neurobiology)

From a physics point of view, the speech signal is based on a change in pressure propagated through the elastic medium 'air', which leads to a specific particle movement. The speech sound is encoded as a complex particle movement in the frequency and intensity domain and is superimposed on nonspeech-related particle movement (noise). Listening to speech-embedded background noise means (1) detecting the particle movement with a sensory system, (2) filtering out nonspeech-related particle movement, and (3) analyzing the remaining speech-related part of the signal and decoding the meaning. Therefore, hearing sounds, especially listening and comprehending language, is a very complex ability, the neurophysiological basis of which is not yet completely understood. This is even more important when the speech signal being focused upon is disturbed by the noise of other speakers, e.g. in the case of listening to a single speech signal during a noisy party. The ability to extract only the relevant parts of the air particle movement related to the speech signal of a certain speaker (so-called 'cocktail party effect') is an enormous achievement of the neuronal structures involved. However, this powerful discrimination of an acoustic signal is not limited to human language processing. For example, seabirds returning to their colony are able to find their straying offspring by differentiating their specific call ('identificators') among the calls of up to some 10,000 other chicks. Although the physiological basis of analyzing complex sounds is phylogenetically old and quite common in animals, this does not diminish the enormous achievement of the human brain to comprehend spoken language.

The speech signal (encoded as the complex movement of air particles) is gathered, filtered, and ampli-fied by the ear, which is divided into three parts: (1) The external ear, which comprises the cartilaginous pinna and a canal (external auditory meatus) leading to a vibratory motion-sensitive membrane (tympanum). As a consequence of the resonance properties of the ear canal, the pinna, and the head, the frequency range of 1,500 to 7,000 Hz (the main speech range) is preferred while other frequencies are reduced (passive filtered). (2) The middle ear, which comprises the very small bones of the ossicle chain (malleus, incus, stapes), transmits the vibratory motions of the tympanic membrane to the much smaller membrane of the 'oval window' of the inner ear. This part amplifies the signal mechanically by increasing the pressure (gain) due to the ratio of the effective areas of the tympanum to the small oval window and by exerting a lever effect of the oscillating ossicles. (3) The inner ear or cochlea, where the motion of the ossicle chain is coupled to the fluids of the cochlear canals, where the sensory cells (hair cells) are located in the organ of Corti of the scala media. Here, transduction from a vibratory motion to an electrical signal takes place. The following stations of the so-called auditory pathway are shown in Figure 2 of **Brain Organization and Auditory Pathways.**

The process of language analysis takes place at an incredible speed. A trained listener can repeat sentences ('shadowing') with a latency of only 220 ms. This is evidence for the assumption of parallel analysis processes during language comprehension. Experiments show that words are almost comprehended after only 100 ms, a time period that comprises only one to two phonemes. The sound signals are mostly discriminated by phylogenetically old neuronal feature detectors. This is why most mammals (e.g. tested in

chinchillas) are able to discriminate different plosives in a choice reaction experiment. After words have been detected and recognized, the cognitive part of language comprehension sets in: analyzing the meaning of the word, analyzing the meaning of phrases and sentences, and analyzing connotative meaning delivered by contextual information, the speaker's prosody, or nonverbal communication signals. More than milliseconds are required to understand language in that sense; it may take up to several seconds before a listener 'gets the point' of a certain utterance (see time scale in Figure 2 of **Brain Organization and Auditory Pathways**). In addition, such complex analyses and judgments of complex utterances are performed by integrative processes of many different brain areas. During language processing, participating brain processes can be investigated with electrophysiological or brain-imaging techniques. At the moment, the minority of experiments are conducted with invasive electrical recording of single or few neurons. Most electrophysiological results are based on noninvasive observation of the synchronous activity of huge groups of neurons (from some 100,000 up to a few million cells).

Modern electrophysiological and brain-imaging techniques such as electroencephalography (EEG), functional magnetic resonance imaging (fMRI), or positron emission tomography (PET) have led to a new empirical basis for the investigation of cognition (*see* **Neurolinguistics**). While event-related fMRI was introduced only a few years ago, other techniques had already provided a substantial number of results. For example, for more than two decades the use of event-related potential (ERP) techniques has yielded new insights into the physiological processes underlying language. The language-related wave form of the scalp EEG first described by Marta Kutas and Steven A. Hillyard in 1980 was the so-called N400, a large negative ERP component related to semantic analysis of sentence processing. For example, the final word in '*The pizza was too hot to cry*' elicits a huge bilateral negativity at posterior electrode positions between 250 and 600 ms after word onset, with a peak latency of about 400 ms, whereas the final word in the sentence '*The pizza was too hot to eat*' does not. However, a sentence with an intermediate semantic anomaly such as '*The pizza was too hot to drink*' would elicit an N400 with a smaller amplitude, since the peak amplitude corresponds to the extent of the semantic anomaly. In contrast, a sentence with a syntactic anomaly like '*Turtles will spit out things they does not like to eat*' would elicit a positive component (P600), peaking about 600 ms after the onset of the word '*does*'. In recent years, the N400 has been shown for many languages, including American Sign Language (ASL). An additional language-related ERP component is the N280, which can be seen as an indicator for processing closed-class words (function words such as articles, conjunctions, or prepositions) in contrast to open-class words (content words such as verbs, nouns, and adjectives).

References

Binder, Jeffrey R., J.A. Frost, Thomas A. Hammeke, R.W Cox, and Thomas E. Prieto. 1997. Human brain language areas identified by functional magnetic resonance imaging. The Journal of Neuroscience 17. 353–62.

Brown, Colin M., and Peter Hagoort (eds.) 1999. The neurocognition of language. Oxford: Oxford University Press.

Fitch, R. Holly, Steve Miller, and Paula Tallal. 1997. Neurobiology of speech perception. Annual Review of Neuroscience 20. 331–53.

Gazzaniga, Michael S. (ed.) 2000. The new cognitive neurosciences, 2nd edition. Cambridge, MA: MIT Press.

Katz, Jack, Nancy A. Stecker, and Donald Henderson (eds.) 1992. Central auditory processing: a transdisciplinary view. St. Louis: Mosby.

Kutas, Marta. 1997. Views on how the electrical activity that the brain generates reflects the functions of different language structures. Psychophysiology 34. 383–98.

Müller, Horst M., Jonathan W. King, and Marta Kutas. 1997. Event related potentials elicited by spoken relative clauses. Cognitive Brain Research 5. 193–203.

Ojemann, George A. 1991. Cortical organization of language. The Journal of Neuroscience 11. 2281–7.

Pulvermüller, Friedemann. 2000. Words in the brain's language. Behavioral and Brain Sciences 22. 253–336.

HORST M. MÜLLER

Speech Production

Research on speech production studies the processes and representations that are involved in producing an utterance, from generating the concept to be expressed, to the motor movements of the articulatory organs. Although language production has historically been the subject of less research than comprehension, recent

years have seen increased interest in the question of how speakers are able to produce complex and well-formed speech in real time. This increased interest has been accompanied by important methodological developments. Until relatively recently, language production was mainly studied through observational methods, by which researchers collected corpora of naturally occurring spontaneous speech, particularly speech errors, and used these to develop theories of the processes that underlie speech production. Observational methods obviously present methodological difficulties, such as observer bias, infrequency of some phenomena, and the extent to which any corpus is representative of the language as a whole. Some of these problems have been solved with the advent of powerful computers that can collect and process large quantities of natural language. However, methodological developments have also made it possible to study the production of both well- and ill-formed utterances under rigorously controlled laboratory conditions. Experimental manipulations now allow such diverse possibilities as inducing the production of particular syntactic structures or speech errors of interest, as well as investigating the link between how we visually perceive the world and how we describe it.

There is perhaps surprising consensus concerning the overall architecture of the language production system. The model has been articulated in more detail by psycholinguists such as Willem Levelt on the basis of recent experimental research; but in essential respects the currently accepted architecture of the language production system remains largely unchanged from that proposed by Merrill Garrett in the 1970s on the basis of speech error data. According to Willem Levelt's account, production comprises three main stages: conceptualization (generating the prelinguistic message), formulation (translating the message into linguistic form), and articulation (executing the motor programs that realize the linguistic representation). Each has its own characteristic input and output representations. Another component, the self-monitor, is assumed to check for errors or other problems at each level of processing.

During conceptualization, the speaker conceives an intention to express, and selects relevant information. The representations implicated here are assumed to be propositional structures, i.e. prelinguistic representations that involve expressions (or constituent parts of those expressions) that can have truth values. They are also assumed to specify the speaker's perspective, the information structure of the sentence, and any features that are encoded by the grammar (e.g. tense information) of that language. Because languages differ in terms of which semantic features are grammatically

expressed, the features expressed in the pre-linguistic message that is the output of conceptualization vary across languages.

During the next stage, formulation, the speaker converts the prelinguistic message into linguistic form. Formulation is subdivided into stages of grammatical encoding and morpho-phonological encoding. Each stage involves building abstract structure, and retrieving content to fill that structure. Grammatical encoding is concerned with developing syntactic structure. The speaker must select a syntactic structure that can express the intended meaning, and congruent lexical content. Research suggests that grammatical encoding is a two-step process of functional processing followed by positional processing. In functional processing, the speaker retrieves lexical concepts (the semantic content of a lexical entry) and assigns functional relations (subject, direct object, indirect object, etc.) to them. This is the stage at which at least some agreement relations are computed. Functional level representations are unspecified for linear order. Next, in positional processing, the speaker generates a syntactic frame and associates it with the lemma, or syntactic content, of each relevant lexical entry. This syntactic frame specifies in which linear order the individual elements are to appear, i.e. each phrase appears in the position that it will occupy in the final utterance.

The output of grammatical encoding forms the input to morpho-phonological encoding, where the speaker translates the syntactic frame and its content into appropriate articulatory gestures. This involves a number of intermediate stages. First, each word's morphological makeup (involving the relevant word forms) must be retrieved. Next, the individual sounds, syllables, and intonational properties of these morphemes must be specified. Finally, a process of phonetic encoding maps from phonological words to gestural scores, i.e. a pronunciation plan is generated.

The last stage in speech production is articulation, where the speaker executes the abstract gestural scores. This involves the storage of phonetic plans in an articulatory buffer, their retrieval as needed, unpacking the hierarchy of motor programs, and issuing instructions to the neuromotor circuits that control the various articulatory muscles.

This overall architecture is widely accepted, but uncertainty remains concerning the details of the production processes. One question of particular interest is how activation passes between different elements of the production system. Some researchers believe that activation can only flow forward, with later levels of processing (e.g. morpho-phonological processing) being unable to affect earlier levels (e.g. functional

processing). Others believe that activation can flow backward as well as forward. Similarly, it is not clear whether processing is discrete, so that processing must be completed at each level before processing can begin at the next level, or cascaded, so that processing of later levels can begin while processing is still under way at an earlier level. It is also unknown as to how far the incrementality of speech production extends: elements appear to be processed as they become available, but the minimal unit of processing at each level remains unclear. More glaring gaps in our understanding of speech production concern the nature of conceptualization, and the coordination of syntactic structure-building, on the one hand, and lexical retrieval on the other.

References

Bock, Kathryn. 1996. Language production: methods and methodologies. Psychonomic Bulletin & Review 3. 395–421.

Bock, Kathryn, and Willem Levelt. 1994. Language production. Grammatical encoding. Handbook of psycholinguistics, ed. by M.A. Gernsbacher. London: Academic Press.

Dell, Gary. 1986. A spreading activation theory of retrieval in sentence production. Psychological Review 93. 283–321.

Garrett, Merrill. 1980. Levels of processing in sentence production. Language production: Vol. 1. Speech and talk, ed. by B. Butterworth. London: Academic Press.

Levelt, Willem. 1989. Speaking: from intention to articulation. Cambridge, MA: Bradfor Books.

Vigliocco, Gabriella, and Robert Hartsuiker. 2002. The interplay of meaning, sound, and syntax in sentence production. Psychological Bulletin 128. 442–72.

HOLLY BRANIGAN AND ROBERT J. HARTSUIKER

Speech Production: Neurobiology

It takes only about two seconds for speakers to compute the correct meaning, grammatical, and sound structures to encode their ideas into speech. How does the brain accomplish this encoding so fast? Which parts of the brain participate in each level of linguistic processing? Neurolinguists have attempted to answer these questions by studying the speech of stroke and cerebral trauma victims who have language impairments called 'aphasias.' These researchers tried to localize, or identify, the regions of the brain responsible for different linguistic processes, such as grammatical encoding and word comprehension. However, continuing research has shown that even if a part of the brain is necessary for a certain aspect of linguistic processing, it is unlikely that this is the only part of the brain involved in that particular linguistic function. Furthermore, studies have shown that brain damage in different parts of the brain can lead to similar language impairments, whereas the inverse is also true: the same locus of brain damage may lead to different language impairments in different people. Knowledge about the contribution of various brain regions to language use is increasing daily because of the development of noninvasive technology such as event-related potentials and neuroimaging techniques such as positron emission tomography, single-photon emission computed tomography, and functional magnetic resonance imaging. These techniques confirm that many parts of the brain contribute to language processing in all linguistic tasks. Nevertheless, various aspects of speech production have proved to be more vulnerable to damage in and around the Sylvian fissure of the left hemisphere, in the regions indicated in Figure 1.

Articulation suffers when there is damage to Broca's area. Damage in this part of the brain, located just forward of the oral motor cortex, impairs the ability to sequence the movements of articulators to produce the sounds of the language. In Broca's aphasia, speech becomes nonfluent and effortful, as if the speaker must consciously control the movements of the articulators to produce spoken words. In fact, the utterances of speakers with Broca's aphasia frequently consist of only one word at a time, with longer utterances being very rare with little grammatical complexity.

Although all aphasics make some errors in pronunciation, the use of sound-related information is most impaired in conduction aphasia, which is caused by damage near the auditory association area in the temporo-parietal cortex. These aphasics produce words that are off-target by one or two sounds, then repeatedly correct themselves as they attempt to produce the target form, e.g. 'I bought a bus, but, boot, book'. Thus, they are aware that they have produced an incorrect word form, but are not sure, until they hear themselves say it, what the correct sound combination is. Importantly, subjects do not consistently make errors on the same word across trials, indicating that the impairment affects the pronunciation of word forms but that the underlying knowledge of the sound structure of words is preserved. Conduction aphasics also
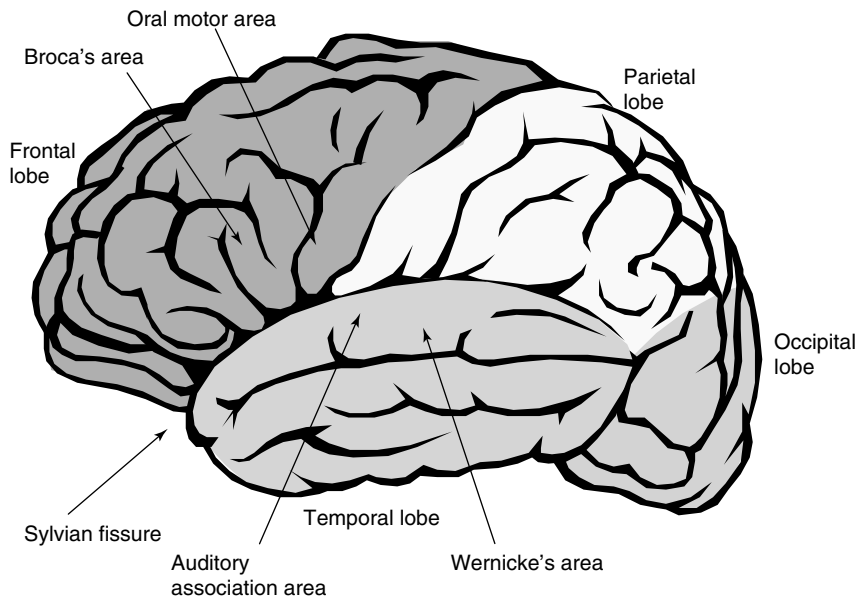
Figure 1. Speech areas in the left hemisphere.

have impaired auditory short-term memory, which interacts with their deficit in pronunciation and results in severe impairments of oral repetition.

Linguistic impairments found in agrammatic aphasia primarily affect words or word parts carrying grammatical information, such as grammatical inflections of verbs and nouns and function words, such as articles and prepositions. Most people with Broca's aphasia are agrammatic, but not all agrammatics have damage in Broca's area. However, all people with agrammatism have damage in or near the Slyvian fissure. In agrammatism, speakers often omit free-standing grammatical words, such as prepositions and articles, but they substitute incorrect inflected forms for correct ones, for example, 'Yesterday dog bark me'. Note that in English it is not clear whether the word form 'bark' has been substituted for 'barked' or whether the past tense inflection (-*ed*) has been omitted. In other words, it is hard to determine from English data whether grammatical inflection is completely unavailable to the aphasic or whether a simpler, uninflected form of a word has been chosen. However, research on aphasic speakers of highly-inflected languages like Russian has shown that the process involves the substitution of one inflected form for another. This means that grammatical information is, in principle, still available to the aphasic, even though the wrong form may be selected.

The connections between the sound form of a word and its meaning are impaired by damage in or near Wernicke's area. This impairment affects both comprehension and production: single-word comprehension is impaired, unrelated words are substituted for appropriate ones (e.g. 'My boss is an envelope'), and patients have no awareness of their own errors. Some

Wernicke's aphasics also produce nonwords, called 'neologisms', as substitutions for real words. Their speech is fluent but full of pronouns and 'empty' words such as *thing, stuff, do, make,* and *get.* Because of unrelated word substitutions, neologisms, and empty speech, it is frequently difficult to understand the speech of Wernicke's aphasics. Consequently, although they produce complex sentences, it is usually impossible to determine whether sentence structures are being used correctly.

The meaning level of language is selectively impaired in Alzheimer's disease and semantic dementia. In these syndromes, word finding is moderately to severely impaired, leading to the production of related words from the same meaning category or more general terms (e.g. saying 'horse' or 'animal' in response to a picture of a deer). These patients score poorly on tasks requiring full knowledge of word meaning, and their spontaneous speech is often empty of content, as in Wernicke's aphasia. These syndromes are characterized by diffuse loss of neurons in the temporal lobe and often throughout the cerebral cortex. Because all aphasics, regardless of the location of cortical damage, produce these same types of semantic errors, it is unlikely that semantic processing is localized to a specific area of the brain.

It is interesting that researchers have not found a particular type of damage that is associated with errors affecting word order or sentence structure, although many researchers see agrammatism as a syntactic impairment. In contrast, all aphasics show varying degrees of impairments of pronunciation, word finding, and understanding of complex sentences. These findings, in conjunction with those of neuroimaging studies, support the assertion that many areas of cerebral cortex contribute to each level of linguistic

processing, although the areas immediately adjacent to the Sylvian fissure seem to hold particular importance for language use.

References

Blumstein, Sheila E. 1995. The neurobiology of language. Speech, language and communication, ed. J.L. Miller and P.D. Eimas. California: Academic Press.

Caplan, David. 1992. Language: structure, processing and disorders. Cambridge, MA: MIT Press.

Garrett, Merrill. 1992. Disorders of lexical selection. Lexical access in speech production, ed. by W.M.J. Levelt, 143–180. Cambridge, MA: Blackwell.

Levelt, W.M.J. 1989. Speaking: from intention to articulation. Cambridge: Cambridge University Press.

MacWhinney, Brian, and Elizabeth Bates (eds.) 1989. The crosslinguistic study of sentence processing. Cambridge: Cambridge University Press.

Obler, L. K., and K. Gjerlow. 1999. Language and the brain. Cambridge: Cambridge University Press.

LORI J.P. ALTMANN

Speech Synthesis

Speech synthesis is the creation of speech sound by means of a machine or a computer. The development of speech synthesis dates back to the end of the eighteenth century when Wolfgang von Kempelen built a speaking machine in Vienna. His apparatus was based on a model of the human vocal tract, with bellows for lungs, a conical resonator that could produce vowel sounds, and constricted portions to create consonantal sounds. Another pioneer was Alexander Graham Bell who, together with his brother, built a talking head with the pharynx, tongue, velum, lips, and teeth made from wood, rubber, and cotton. In the twentieth century, with the invention of electronic circuitry, the vocal tract transfer functions were simulated by (combinations of) sound sources and filters.

Stored Speech

The task of the synthesizer is to create and produce sounds that are recognized as speech. One way to accomplish this is to make use of stored utterances. In general, the quality of prerecorded speech is high, but the vocabulary is fixed and restricted, and is only useful for reproduction (e.g. in cars and aeroplanes). Moreover, despite coding techniques, this synthesis method requires considerable memory capacity for storage. More flexible systems require the conversion of words to a string of phonemes, make use of pronunciation rules (between phonemes and sounds) stored in dictionaries, and add durational and prosodic information. Several text-to-speech synthesis techniques exist, none of which generate speech that sounds completely natural. At present, the two most common (computer) techniques are parametric coding (also called formant synthesis by rule) and concatenative synthesis.

Parametric Coding

This type of synthesizer makes use of the source–filter theory of speech production and strings together several acoustical properties of speech sounds (by means of a set of rules). The sound source is characterized by voiced pulses or noise. Subsequently, filters are used to modify the source, i.e. to produce different resonances or formants of vowel sounds or to pass high-or low-frequency information of consonants. In some systems, several parameters can be altered to convey gender and age. For example, the word 'seek' is generated by first passing noise through a high-pass filter (at about 5,000 Hz), followed by a low first formant (350 Hz) together with a high second formant (2,200 Hz) and high third formant (3,000 Hz) for the vowel /i/, a short silence to simulate pressure buildup before unvoiced plosives and a sharp cutoff sound for /k/. Rules are used to simulate natural timing, stress, and prosody and to create natural connections between the speech sounds.

The advantage of formant-based synthesis is that it requires neither a large storage capacity nor high computational demands. As the acoustic differences between speech sounds can be varied very precisely, this type of synthesis is commonly used for research purposes, for instance, to examine which acoustic properties of speech sounds determine the identity of the speech sound. The disadvantage of this type of synthesis is that it proves to be very difficult to connect the different parameter tracks between adjacent sounds in a natural way. The same speech sound produced in a different context sounds different (e.g. aspirated, unaspirated). Ideally, each different realization of the same speech sound (phoneme) should be represented by the synthesizer. Moreover, speech sounds do not

occur neatly one after the other during production due to coarticulation. Another problem involves the assignment of the correct timing and intonation (cf. 'unit' and 'undo'), and stress differences (cf. 'project' and 'to project'). Although different sets of rules exist to distinguish pronunciations, and to assign syntactic structure (to specify the pronunciation of sound sequences), it remains very difficult to determine the right settings at the right time.

Concatenative Synthesis

Some of the above-mentioned problems can be solved by concatenating pieces of natural recorded speech. For this purpose, diphones and demisyllables are usually used. A diphone is a piece of speech that stretches from the middle of one speech sound to the middle of the following one. They are cut from several recorded utterances, stored, and recombined during synthesis of an utterance. A demisyllable is half a syllable, with the cut in the middle of the vowel. Synthesis of unrestricted English requires approximately 1,000 diphones or demisyllables. Concatenative synthesis is usually highly intelligible and sounds very natural, because the articulatory effects from one phoneme to the next are included. Although the advantage of this type of synthesis is clear, discontinuities can still arise at the boundaries of the diphones or the demisyllables. Moreover, it is not possible to record all existing words spoken with all possible pronunciations. As this type of synthesis also requires considerable storage capacity, voice coding techniques are used to compress the speech sounds.

Application

The best synthesis technique depends on the requirement of the intended application. There is always a trade-off between maximizing speech quality, minimizing memory capacity, and computation time. Speech synthesis is used, among other things, for information retrieval (e.g. banking), for warnings (e.g. in cockpits), in reading machines for the blind, for speech-impaired people, for foreign language learning, in toys, and in speech research. The improvement of speech synthesis lies primarily in a more natural generation of the prosody of speech. Although synthetic speech is highly intelligible, it usually takes half a sentence to know that it is not human. This is often because the listener is distracted by small details, such as timing, voice quality, overarticulation, or unnatural intonation. Whatever the use of synthesized speech, the issues of intelligibility and naturalness will remain challenging in the forthcoming years.

References

Allen, J.A., M.S. Hunnicutt, and D.H. Klatt. 1987. From text to speech: the MITalk system. Cambridge, England: Cambridge University Press.

Bailly, G., and C. Benoit. 1992. Talking machines — theories, models and designs. Amsterdam: North-Holland.

Klatt, D.H. 1980. Software for a cascade/parallel formant synthesizer. Journal of the Acoustical Society of America 67. 971–95.

———. 1987. Review of text-to-speech conversion for English. Journal of the Acoustical Society of America 82. 737–93.

Klatt, D.H., and L.C. Klatt. 1990. Analysis, synthesis and the perception of voice quality variations among female and male talkers. Journal of the Acoustical Society of America 87. 820–57.

Kohler, K.J. 1991. Prosody in speech synthesis: the interplay between basic research and TTS application. Journal of Phonetics 19. 121–38.

Liberman, A.M., F. Ingemann, L. Lisker, P. Delattre, and F.S. Cooper. 1959. Minimal rules for synthesizing speech. Journal of the Acoustical Society of America 31. 1490–9.

Pierrehumbert, J. 1981. Synthezing intonation. Journal of the Acoustical Society of America 70. 985–94.

Pisoni, D.B. 1997. Perception of synthetic speech. Progress in speech synthesis, ed. by J. Van Santen et al., 541–60. New York: Springer-Verlag Inc.

Van Heuven, V., and L.C.W. Pols. 1993. Analysis and synthesis of speech-strategic research towards high-quality text-to-speech generation. Berlin: Mouton de Gruyter.

ASTRID VAN WIERINGEN

Sprachbund

The notion of *Sprachbund* (or linguistic area) has been introduced in linguistic theory by Nikolaj Trubetzkoj in 1928. In his famous proposition XVI, put forward on the occasion of the First International Congress of Linguists at The Hague, Trubetzkoj defines *Sprachbünde* as made up of languages displaying a high degree of similarity in syntactic form, some similarity in the principles that govern morphology, and a great number of common culture words—sometimes also an exterior similarity in the inventory of sounds. In modern linguistic theory, *Sprachbund* usually designates an area characterized by a number of linguistic

features shared by a number of languages either unrelated or from different subgroups of the same family. The following linguistic areas have been identified so far: (i) The Balkans (Greek, Albanian, Serbo-Croatian, Bulgarian, Macedonian, and Rumenian); (ii) South Asia (Indian subcontinent); (iii) Meso-America; (iv) Amazonia; (v) Ethiopia; and (vi) Europe.

Each of the aforementioned linguistic areas displays a number of distinctive features. For instance, the Balkan linguistic area, which is by and large the most famous linguistic area identified so far, is characterized by the following features: (a) formation of future and perfect tense using an auxiliary; (b) replacement of infinitives by finite verb forms; (c) postponed articles; and (d) merging of dative and genitive case forms.

Areal linguistics (i.e. the subfield of linguistics which aims at uncovering linguistic areas) has been facing the problem of the number of features necessary to distinguish between proper *Sprachbünde* and simple contact phenomena among a group of languages. According to a widely held view, there should be genealogical diversity of the languages belonging to a linguistic area, although it cannot be excluded that shared features may be due to language contact even in the case of genetically related languages. Moreover, the similarities that characterize a linguistic area should not be trivial, i.e. they should not be the result of typological developments that take place, under the same circumstances, in all or most of the world's languages. Language typology, being traditionally concerned with the classification of languages based on grammatical features independent of genetic relationship, may provide areal linguistics with some evaluation criteria. Similarly, historical linguistics may help areal linguists to exclude that some postulated areal features are due to genetic inheritance.

In some recent developments of areal linguistics, the notions of areality are considered to be of interest irrespective of whether the phenomena examined can be described in terms of linguistic areas in the traditional sense. This change in scientific perspective has led to a rather composite approach to linguistic diversity that could be labeled as 'areal typology' (Östen Dahl 2001): to the extent that typologically interesting features are not evenly spread geographically but tend to cluster in certain areas, areal typology can be considered as the study of the geographical distribution of linguistic features, rather than the characteristics of individual areas. In this vein, Johanna Nichols' important monograph (1992) has provided a complex model of linguistic diversity, which involves two main areal patterns, namely spread and residual zones, the former being areas in which languages tend to spread and mix quickly and repeatedly and the latter being zones where languages tend to remain undisturbed over long periods.

This perspective has been recently applied to the study of Circum-Baltic languages (Koptjevskaja-Tamm and Wälchli 2002) and Mediterranean languages (Ramat and Stolz 2002). In both the Circum-Baltic and the Mediterranean area, there are several convergence/contact phenomena on the microlevel, that interact with phenomena on the macrolevel. This distinction between microlevel and macrolevel phenomena nicely accounts for the internal heterogeneity of linguistic areas. The claim underlying such studies is that the notion of *Sprachbund* tends to overemphasize the overall macrocontact, which is of course justified in certain areas (as the well-established Balkan linguistic area). In other areas, however, intensive microcontacts superimposed on one another may create the impression of an overall macrocontact among the languages, which has not necessarily being there. The obvious consequence of this assumption is that areal phenomena are worth looking at in their own right, even if we are not faced with linguistic areas.

References

Bisang, Walter. 1998. Areal typology and grammaticalization: processes of grammaticalization based on nouns and verbs in East and mainland South East Asian languages. Studies in Language 20.

Campbell, Lyle, Terrence Kaufman, and Thomas C. Smith-Stark. 1986. Meso-America as a linguistic area. Language 62.

Dahl, Östen. 2001. Principles of areal typology. Language typology and language universals, ed. by Martin Haspelmath, Ekkehard König, Wulf Österreicher, and Wolfgang Raible. Berlin–New York: Mouton de Gruyter.

Ferguson, Charles A. 1976. The Ethiopian language area. Languages in Ethiopia, ed. by M.L. Bender, J.D. Bowen, and Charles A. Ferguson. London: Oxford University Press.

Haspelmath, Martin. 1998. How young is standard average European. Language Sciences 20.

Koptjevskaja-Tamm, Maria, and Bernhard Wälchli. 2002. An areal typological approach to the Circum-Baltic languages. The Circum Baltic languages. Their typology and contacts, ed. by Östen Dahl and Maria Koptjevskaja-Tamm. Amsterdam–Philadelphia: Benjamins.

Masica, Colin P. 1976. Defining a linguistic area. South Asia. Chicago: University of Chicago Press.

Nichols, Johanna. 1992. Linguistic diversity in space and time. Chicago: University of Chicago Press.

Ramat, Paolo. 1998. Typological comparison and linguistic areas: some introductory remarks. Language Sciences 20.

Ramat, Paolo, and Thomas Stolz (eds.) 2002. Mediterranean languages. Papers from the MEDTYP Workshop, Tirrenia, June 2000. Bochum: Universitätsverlag Dr. N. Brockmeyer.

Sandfeld, K. 1930. Linguistique Balcanique. Problèmes et Résultats. Paris: Klincksieck.

Trubetzkoj, Nikolaj S. 1930. Proposition 16. Actes du Premier Congrès International de Linguistes à la Haye, du 10–15 Avril 1928. Leiden: A.W. Sijthoff.

ANDREA SANSÒ

See also **Balkans; Historical Linguistics; Language: Contact—Overview; Trubetzkoy, Nikolai Sergeyevich; Typology**

Sranan

Sranan (or Sranan Tongo, lit. 'language of Suriname') is the name of the English-lexicon creole language that has been used in Suriname (in the northeast of South America) since the late seventeenth century. In the past, it has also been referred to as Negro-English, Nengre, or, in a rather derogatory manner, Taki-Taki. It is spoken by some 350,000 people in Suriname, French Guyana, and the Netherlands, either as a first or as a second language. It is the native language of most Surinamese people of African descent, while it serves as an interethnic lingua franca between the other ethnic groups, which include Amerindians, Indians, Javanese, and Chinese. Although it does not have any official status—the only official language in Suriname is Dutch—it is being used more and more in formal contexts, such as education, the media, politics, and public information. Apart from this, there is also a flowering literature in Sranan, especially poetry.

The reason why an English-lexicon creole is spoken in a country that was a Dutch colony throughout most of its existence is purely historical. Before it became a Dutch possession in 1667, Suriname had been an English colony for 17 years, and it is generally assumed that the basis of the Sranan lexicon stems from that period. However, although most of the English had left the colony by 1680, this did not put a definitive stop to the presence of the English in Suriname. Apart from the fact that some of the 'old' English planters returned to Suriname later, new English planters settled there as well. This means that the window of opportunity for Sranan to acquire an English-based lexicon remained open for a longer time than the 30-year period between 1651 and 1680.

Apart from the English element, the Sranan lexicon contains several other layers as well. First, due to the fact that many of the planters in the early period were Portuguese-speaking Jews, a number of Portuguese-derived words have been incorporated into the language. Second, many items from local domains, such as flora and fauna, have Amerindian names, mainly from Arawak and Carib. Third, a fair number of words have been adopted from some of the African languages spoken by the slaves, especially Gbe (a Kwa language, spoken in Ghana, Togo, and Benin), Akan (another Kwa language, spoken in Ghana), and Kikongo (a Bantu language, spoken in Congo-Brazaville and Congo-Kinshasa). Finally, ever since the beginning of Dutch rule, a large number of Dutch-derived words have been borrowed, a process that continues to the present day.

Sranan is somewhat unusual among creoles in general in that it is relatively well documented in its early stages of development. Many written documents—either in printed or manuscript form—are available from the early eighteenth century onward. This is largely due to the work of the Moravian Brethren, a German missionary organization that was very active in Suriname and that produced a large number of religious texts in Sranan as well as a number of invaluable descriptive works such as dictionaries and grammars. Apart from this, there are several printed language primers and dictionaries that were authored by members of the colonial elite. The availability of these early documents has enabled linguists to carry out detailed diachronic research on Sranan, especially on the development of its phonology and syntax (Arends 2002).

As a creole language, Sranan is the product of a process of language contact involving a number of different languages from different language families. Historical research has shown that while the African population consisted of many different ethnolinguistic groups, the majority among them belonged to one of the three larger groups, who spoke either Gbe, Kikongo, or Akan (Arends 1995). This means that the major linguistic input into the creolization process, as far as the African side is concerned, came from these languages. The predominant role of Gbe, Kikongo, and Akan is confirmed by the fact that the African element of the Sranan lexicon is overwhelmingly derived from these three languages. This is further supported by research into other domains such as phonology, lexical semantics, and morphosyntax.

As far as the European population is concerned, it is important to realize that although Suriname was a Dutch colony from 1667 onwards the Dutch were never a majority among the European population. From the late seventeenth to the early nineteenth century, it was the Portuguese Jews who were the numerically most important group among the European population, which, apart from the Dutch, included Germans, French, Scandinavians, and others. As a result, Dutch did not become a majority language among Suriname's Europeans until well into the nineteenth century. This may explain why Sranan was widely used by the Europeans among themselves, even though 'officially' they held it in low esteem.

Below, the major features of each of the linguistic subsystems will be briefly discussed (largely based on Bruyn 2002; see also Adamson and Smith 1995). It

will be followed by a few remarks on the 'verbal arts', an important activity in traditionally oral languages such as Sranan.

Lexicon. About three quarters of the basic vocabulary items are derived from English, while most of the remainder is from Dutch. As for nonbasic vocabulary, this consists mainly of words derived from Dutch, although some items have been taken from other sources, such as Portuguese, Amerindian languages (Arawak, Carib), and African languages. Some examples are *katibo* 'slave' (< Portuguese), *kruyara* 'dug out canoe' (< Arawak), *awara* 'palm species' (< Carib), *agama* 'lizard species' (< Gbe), *pinda* 'peanut' (< Kikongo), *gongosa* 'gossip' (< Akan), and *nyan* 'eat' (< Wolof). Although most English function words were not retained in Sranan, many functional elements are still expressed by English-derived words. This is a result of the fact that English-derived content words from the Sranan lexicon were recruited to fulfill that role. Such a process of grammaticalization results in a content word becoming more grammatical, often being reduced both in its semantic content as well as its phonetic substance. For example, while the English article 'the' was not adopted in Sranan, two definite articles *a* (singular) and *den* (plural) developed out of the demonstrative *dati* 'that' (< English 'that') and the personal pronoun *den* 'they' (< English 'them')respectively. Similarly, while the English indefinite article 'a' was not retained in Sranan, a new indefinite article *wan* 'a' developed, based on the numeral *wan* 'one.' As for the words that were retained from English, their meaning is not necessarily the same as that of its English etymon. This appears, for example, from *anu* (< 'hand'), which means both 'hand' and 'arm,' and *futu* (< 'foot'), which means both 'foot' and 'leg.' Semantic shifts such as these often have their origin in one or more of the African languages spoken by the slaves, many of which have only one word for 'hand+arm' and for 'foot+leg,' respectively. African origins are also responsible for the existence of a word class known as ideophones, words whose only function is to intensify or specify the meaning of another word with which they occur in a fixed combination. For example, the ideophone *fáán*, used to intensify the meaning of the adjective *weti* 'white,' is probably from Gbe. An example is *a weti so fáán* 'he is so very white' (lit. he white so IDEOPHONE).

Phonology. In phonology, there is a clear tendency toward an open syllable structure, which becomes clear from the tendency to add an extra vowel to English-derived words ending in a consonant, such as *udu* 'wood' (< 'wood') and *waka* 'walk' (< 'walk'), or to insert a vowel into some consonant clusters.

Word-final nasals are velarized, something that is reflected in the spelling, as in <Sranan>, which is pronounced [Sranang].

Morphology. The four morphological processes operative in Sranan are conversion, compounding, suffixation, and reduplication. Conversion (also known as multifunctionality or zero derivation) refers to the derivation of a word, e.g. a verb, from another word, e.g. a noun, without any overt change in form. For example, from the adjective *ebi* 'heavy', both a noun *ebi* 'weight' and a verb *ebi* 'to weigh' have been derived in this way. Compounding is a rather common process in Sranan, especially when both elements are nouns, as in *man-pikin* 'son' (lit. 'man child') and *uman-piki* 'daughter' (lit. 'woman child'). One of the few cases of inflection is the use of the noun *man* 'man' as an agentive suffix as in *siki-man* 'sick person' (lit. 'sick man') and *bere-man* 'pregnant woman' (lit. 'belly man'). Finally, reduplication, the creation of a new word by (partially) doubling an already existing word, is quite common. It can be used to create new words belonging to a category different from the base word, as appears from *sisibi* 'broom' (< *sibi* 'to sweep').

Syntax. Sranan is a strict S(ubject)-V(erb)-O(bject) language with a strong tendency toward an isolating morphology. The latter appears from the fact that tense (T), mood (M), and aspect (A) are expressed through independent particles, which are preposed to the verb, rather than through inflection. The TMA system being far too complex to be discussed in any detail, suffice it to say that the distinction between stative verbs ('love') and nonstative verbs ('eat') is of paramount importance for the functioning of the system. For example, while the bare form of a nonstative verb indicates past tense, a bare stative verb indicates present; to indicate present tense, a nonstative verb has to be preceded by the particle *e*. Compare the following examples: *mi nyan fisi* 'I ate fish'; *mi lobi fisi* 'I love fish'; and *mi e nyan fisi* 'I'm eating fish'. While the distinction between stative and nonstative verbs also plays a role in the use of the particle *ben*, other factors, such as discourse structure, come into play here as well, rendering this area of Sranan syntax too complex to be treated any further here. Like many other creoles, Sranan has two copula forms, one (*de*) for location, possession, and existence, and the other (*a*) for nominal predication (although *de* is sometimes used here as well). Adjectival predicates are treated on par with verbal predicates, i.e. they normally follow the subject without an overt copula being inserted in between, as in *yu futu bigi* 'your feet are big' (lit. 'your feet big'). To express intensity or contrast, both verbal and adjectival predicates may be clefted, with a copy of the

predicate left behind, as in *na bigi yu futu bigi* 'Your feet are really big' (lit. 'is big your feet big'). Finally, a syntactic phenomenon known in many creoles is the serial verb construction, where one subject is connected with two or more main verbs, which together form one semantic unit, as in *Rudy ben tyari den buku kon na ini a oso* 'Rudy has brought the books into the house' (lit. 'Rudy has carried the books come at in the house'). In this sentence, the meaning of what is expressed by the preposition 'to' in English is expressed by the verb *kon* 'come,' which forms a series with the verb *tyari* 'carry'. Sranan has a wide variety of different types of serial verb constructions, for the expression of direction, location, instrumental, dative, benefactive, causative, comparative, completion, and complementation. Since both predicate clefting and serial verbs are common features of many West African languages, it seems justified to interpret the occurrence of these constructions in Sranan as retentions from the African languages spoken by the slaves.

Verbal arts. The domain of language use known as the 'verbal arts' includes such activities as story telling and the performance of song and drama. Probably the best known genre is the so-called *Anansi tori*, a type of story named after the trickster-spider Anansi, but including other types of folk tales as well. Although the canonical context for telling *Anansi tori* is at funeral wakes, they may be told on other occasions as well. Both the content and the performative structure of these tales have their roots in West Africa. The basic pattern is the call-and-response structure known from many African-American oral genres, with the story-teller being interrupted by members of the audience punctu-

ating the story with remarks, songs, or even entire 'sub-stories' of their own. The importance of songs, e.g. as an emotional outlet for the slaves, is already apparent from early sources, where reference is made to a social activity known as *pree* 'play' in which dance played an important role. A similar role was played by various kinds of drama, whose origin similarly lies in the plantation period and which continue to be performed to the present day (for splendid collections of Sranan oral literature, see Herskovits and Herskovits (1936) and Voorhoeve and Lichtveld (1975)).

References

Adamson, Lilian, and Norval Smith. 1995. Sranan. Pidgins and Creoles: an introduction, ed. by Jacques Arends, Pieter Muysken, and Norval Smith, 219–32. Amsterdam and Philadelphia, PA: Benjamins.

Arends, Jacques. 1995. Demographic factors in the formation of Sranan, The early stages of creolization, ed. by Jacques Arends, 233–85. Amsterdam and Philadelphia, PA: Benjamins.

———. 2002. Young languages, old texts: early documents in the Suriname Creoles. In Carlin and Arends, 183–205.

Bruyn, Adrienne. 2002. The structure of the Surinamese creoles. In Carlin and Arends. 153–82.

Carlin, Eithne, and Jacques Arends (eds.) 2002. The languages of Suriname. Leiden: KITLV Press.

Herskovits, Melville, and Jean Herskovits. 1936. Suriname folk-lore. New York: Columbia University Press.

Voorhoeve, Jan, and Ursy Lichtveld. (eds.) 1975. Creole drum: an anthology of Creole literature in Surinam. New Haven and London: Yale University Press.

JACQUES ARENDS

See also **Aspect; Function Words; Grammaticalization; Mood; Pidgins and Creoles; Saramaccan; Serial Verb Constructions**

Sri Lanka Portuguese

On May 18, 1498, the Portuguese explorer Vasco da Gama, having discovered a sea route around the Cape of Good Hope, arrived off the coast of Calicut in southwestern India. This event marked the beginning of the first large-scale European presence in South Asia since the expedition of Alexander the Great. As a result of Portuguese contacts, creole languages with Portuguese-based lexicons developed in various coastal communities on the subcontinent, such as Daman, Chaul, Nagapatnam, and Cochin. Collectively known as Indo-Portuguese, these languages all adapted the majority of

their word stock from Portuguese, but developed innovative grammars that combined structural features from Portuguese and local languages with new elements. One of the most important of these creoles is the one still spoken in Sri Lanka (formerly Ceylon). Sri Lanka Portuguese had a longer reign as a lingua franca (language of intergroup communication) than its sister creoles in India and possessed a more copious literature. In addition, it is rivaled for current vitality only by the northern dialects spoken in Korlai, near Bombay (see Clements 1996) and Daman. Sri Lanka Portuguese

is also the most fully described variety of Indo-Portuguese, thanks to the work of Wesleyan missionaries in the early 1800s and to the description published in 1900 by a Goan jesuit priest, Sebastião Rodolfo Dalgado (see Further Reading below). Because of its longevity and voluminous literature, the language is important not only for the study of Indo-Portuguese but also for investigation into the general processes involved in language contact.

Generally, only one language other than Portuguese was spoken in the settings in which the various varieties of Indo-Portuguese developed. This fact sets Indo-Portuguese apart from many other creoles (particularly those labeled 'plantation creoles'), which developed in multilingual settings. In Sri Lanka, two major indigenous languages are spoken: Sinhala, an Indo-Aryan language, and Tamil, a language of the Dravidian family. The initial development of Sri Lanka Portuguese took place among Sinhala speakers along the southwestern coast of the island.

Contacts between Portugal and Sri Lanka began in the early sixteenth century, and these eventually led to the launch of an expedition from Goa in 1517 to establish a fortified trading post at Colombo. Over the following 100 years, the Portuguese extended their control around the entire coast, leaving the independent Kingdom of Kandy in the central highlands. Throughout the area under Portuguese dominion, communities that used Sri Lanka Portuguese as their first language arose as a result of unions between Portuguese soldiers and local women.

Beginning with the union of the Spanish and Portuguese crowns in 1580, events in Europe weakened Portugal's ability to control its vast overseas possessions, which then began to fall prey to the rising power of Holland. The Dutch began cultivating relations with Kandy in 1602, and in 1632 the newly enthroned Raja Sinha II asked for their assistance in expelling the Portuguese. After a long period of conflict, the last Portuguese strongholds capitulated in 1658. From this point on, Sri Lanka Portuguese was isolated from Standard Portuguese. Thus, unlike many creoles, it was not later influenced by the standard language from which it derived its word stock.

Interestingly, because Sri Lanka Portuguese was already established as a lingua franca for communicating with local people, the Dutch continued to use it throughout their rule, and it even became the home language of the families that the Dutch started locally. Dutch and Sri Lanka Portuguese thus came to stand in a relationship known as diglossia, with the former being used as a 'high' (i.e. socially prestigious) language in administration, the courts, and the Dutch Reformed Church, and the latter as a 'low' language for informal face-to-face communication.

The Dutch in their turn were supplanted in 1898 by the British, who, with their victory over Kandy in 1815, completed their dominion over the entire island and held it until Sri Lanka regained its independence in 1948. The British did not copy the Dutch in establishing local families and thus never adopted the creole themselves, but continued to use it as an expedient lingua franca until its functions could be taken over by English. This replacement appears to have occurred around the middle of the nineteenth century. Certainly by 1875, Sri Lanka Portuguese is reported by the Portuguese creolist, Adolpho Coelho, to be 'almost exclusively the language of the descendants of the Portuguese and Dutch'.

Following Dalgado's description of Sri Lanka Portuguese, published in 1900, we heard little further until it was 'rediscovered' in the 1960s by the late Dr. D.E. Hettiaratchi of the University of Sri Lanka, Peradeniya, and editor of the Sinhala Encyclopedia. Currently, the creole is known to be spoken only by 'Portuguese Burghers' in the east-coast towns of Trincomalee and Batticaloa in the predominantly Tamil-speaking area of the island. Due to the modern civil strife in Sri Lanka, recent figures on the number of speakers are not available, but clearly the language is moribund. All Sri Lanka Portuguese speakers are bilingual in Tamil, with some also speaking Sinhala and/or English. In many families, Tamil has supplanted the creole as a home language. This is partly due to marriage with non-creole speakers and partly due to the fact that men must often seek work elsewhere on the island, where they are in a purely Tamil or Sinhala-speaking milieu.

The sounds of sixteenth-century Middle Portuguese are generally found in Sri Lanka Portuguese, although their pronunciation has taken on a local flavor. For example, Sri Lanka Portuguese /v/ is pronounced [ʋ], a sound found in languages throughout the region, and that can be heard in the typical South Asian pronunciation of *vine* or *wine*. On the other hand, Sri Lanka Portuguese sometimes preserves a Middle Portuguese pronunciation now lost in standard varieties of Portuguese; thus, the first consonants of *cegaa* ('arrive;' Standard Portuguese *chegar*) and *juustu* ('correct;' Standard Portuguese *justo*) are pronounced [tʃ] and [dʒ], while in both Portugal and Brazil they have changed to [ʃ] and [ʒ]. Some Middle Portuguese sounds have also disappeared; for example, *x* (pronounced [ʃ]) has been replaced by /s/, as seen in *basu* ('below;' Standard Portuguese *baixo*). More importantly, the entire set of nasalized vowels has disappeared, although an *m* usually appears as evidence of former nasalization. This can be seen in examples such as *paam* ('bread'; Standard Portuguese *pão*) and *amiyaam* ('tomorrow'; Standard Portuguese *amanhã*). Finally, Sri Lanka Portuguese has developed some new sounds.

For example, as in both Sinhala and Tamil, the creole now has a contrast between long and short vowels, which can be seen in pairs such as: *ooy* ('eye'; Standard Portuguese *olho*) vs. *oy* ('today'; Standard Portuguese *hoje*), and *triisti* ('sad'; Standard Portuguese *triste*) vs. *isti* ('this'; Standard Portuguese *este*).

In its grammar, even more than in its phonology, Sri Lanka Portuguese has a predominantly South Asian rather than European character. This can be seen in such features as word order, noun and verb categories, and complex sentence structures involving relative clauses and the quotative construction (explained below). Moreover, Standard Portuguese characteristics such as noun gender and subject–verb agreement are absent from the creole.

The order of the main elements in a Sri Lanka Portuguese sentence is subject–object–verb; for example, *eev eli-pa diñeeru jaa-daa* ('I him money gave' or 'I gave him the money'; Standard Portuguese *Dei o dinheiro para/a ele* or *Lhe dei o dinheiro*). Adjectives and other noun modifiers precede their nouns: *nosa noov ravkiiña* ('our new violin'; Standard Portuguese *nossa rauquinha nova*). Postpositions rather than prepositions are found: *ælmææra ñiva* ('wardrobe on' or 'on top of the wardrobe'; Standard Portuguese *acima do armário*). In all these word-order characteristics, Sri Lanka Portuguese follows the model of Sinhala and Tamil rather than that of Standard Portuguese.

The role that nouns play within a sentence structure is indicated by case suffixes and postpositions, again, as in Sinhala and Tamil. The dialect spoken in Batticaloa has four case suffixes; for example, *ɔɔmi* 'man' has the following forms: *ɔɔmi-pa* (object of verb, recipient, 'to the man'), *ɔɔmi-su* (possessor, 'man's'), *ɔɔmi-ntu* (location, 'in the man'), and *ɔɔmi-ntaa* (person spoken to, person holding something, 'with the man').

The verbal system is interesting because, unlike Sinhala and Tamil, Sri Lanka Portuguese uses prefixes to indicate tense; thus, for the verb *kuziñaa* 'cook', the following forms are found: *ta-kuziñaa* (present), *jaa-kuziñaa* (past), and *lo-kuziñaa* (future). These structures are also quite unlike Standard Portuguese: compare *cosinha* '(he or she) cooks', *cosinhou* '(he or she) cooked', *cosinhará* '(he or she) will cook'. Other prefixes indicate completion, obligation, negation, and conditionality, as in *mes-kuziñaa* 'must cook', and *naa-kuziñaa* 'won't cook'. These prefixes were developed from a variety of Standard Portuguese sources: auxiliary verbs (*ta-* from *está* 'be'), adverbs (*jaa* from *ja* 'already', *lo-* from *logo* 'soon'), adjectives (*mes-* from *mister* 'necessary'), etc. Some auxiliary verbs appear before the main verb, as in Standard Portuguese; others appear after the verb, as in Sinhala and Tamil; some can appear in both positions. Examples are: *kera vii* ('intends to come; Standard

Portuguese *quer vir*); *falaa tiña* ('had told;' Standard Portuguese. *tinha falado*); *naa-pooy botaa* or *pa-botaa naa-poy* ('can't put;' Standard Portuguese *não pode botar*). The categories that these verbal structures represent are generally those of Tamil and Sinhala rather than Standard Portuguese; the completion category is one example, but many others arise. Moreover, Standard Portuguese verbal categories that are not found in Sinhala and Tamil, such as the subjunctive, are also not found in the creole.

Complex sentence structures are also generally modeled on those of Sinhala and Tamil. Relative clauses, for example, always come before the noun they modify and do not use relative pronouns. An example is shown below with Sinhala and Tamil equivalents:

SLP	*botus*	*diñeeru*	*ja-daa*	*pesaam*
Sinhala	*mahattea*	*salli*	*diy-e*	*miniha*
Tamil	*niinka*	*calli-ya*	*kuʈu-tt-a*	*aaḷ*
Gloss	you	money	gave	person

('The person to whom you gave the money'; Standard Portuguese *a pessoa a quem o Senhor dou o dinheiro*).

Another construction typical of Dravidian languages and shared by Sinhala is the so-called 'quotative construction', which marks the object of an overt or implied verb of thinking, naming, speaking, etc. The quoted material is followed by the quotative particle *falaa* or *falaatu* in Sri Lanka Portuguese, *kiyələ* in Sinhala, *enʈu* in Tamil. The quotative particle derives in all three languages from a verb meaning 'say'. In the example below, both the verb 'say' and the quotative particle appear:

SLP	*eli*	*jaa-falaa*	*eev*	*ja-kaa-falaa*	*falaatu*
Sinhala	*eyaa*	*kiwwe*	*mamə*	*kiww-e*	*kiyələ*
Tamil	*avan*	*connaan*	*naan*	*coll-i-ṯ-ṯ-an*	*enʈu*
gloss	he	said	I	have told	QUOTE

('He said "I have told [him]"' or 'He said he had told [you]'; Standard Portuguese *Ele disse que 'eu [lhe] falei'.* or *Ele disse que [lhe] tinha falado.*)

The vast majority of Sri Lanka Portuguese words are Portuguese in origin, but many other items are borrowed from Dutch, English, Sinhala, and Tamil, such as *gɔrgal* 'throat' (Du. gorgel), *ḍɔkʈa* 'doctor' (Eng.), *suura* 'toddy, a drink made from fermented palm sap' (Si. *surə*), *neeli* 'paddy, unhusked rice' (Ta. *nellu*). Occasionally, Sri Lanka Portuguese preserves a Middle Portuguese word no longer current in Standard Portuguese, such as *izmaleeru* ('beggar'; Standard Portuguese *esmoleiro*, archaic).

Literary Sri Lanka Portuguese is largely a nineteenth-century phenomenon. Kenneth David Jackson's book *Sing without shame* demonstrates that an oral tradition existed, which, because of its ties with the

traditions of other Portuguese-speaking communities, probably originated during Portuguese rule on the island. The earliest extant printed records of the creole are the work of Wesleyan missionaries, who made extensive use of Indo-Portuguese in both India and Sri Lanka (see Fox 1819.) Either through ignorance or because of a desire to dignify the low-prestige creole, the missionaries dressed Sri Lanka Portuguese in the trappings of the more prestigious standard language to which it is related, producing a hybrid variety combining Standard Portuguese spelling and generally Standard Portuguese and English grammar with some aspects of creole word structure, particularly in the use of verb prefixes.

The fact that we have reason to be skeptical of nineteenth-century Sri Lanka Portuguese texts should make us equally skeptical of nineteenth-century creole texts in general, and of those purporting to represent other varieties of Indo-Portuguese in particular.

References

Clements, J. Clancy. 1966. The genesis of a language: the formation and development of Korlai Portuguese. Amsterdam, Philadelphia: John Benjamins.

Coelho, Adolpho F. 1967. Os dialectos românicos ou neolatinos na África, Ásia e América. Bolletim da Sociedade de Geografia de Lisboa, 2:3 (1880), 3:8 (1882), 6:12 (1886), reprinted in Estudos linguisticos crioulos, ed. by Jorge Morais-Barbosa. Lisbon: Academia Internacional de Cultura Portuguesa.

Dalgado, Sebastião Rodolfo. 1900. Dialecto indo-português de Ceylão, Lisbon: Imprensa Nacional; reissued with introduction by Ian R. Smith, Lisbon: Comissão Nacional para as Comemorações dos Descobrimentos Portuguêses, 1998.

Fox, William Buckley. 1819. A dictionary in the Ceylon-Portuguese, Singhalese and English languages ... to which is prefixed a compendium of the Ceylon-Portuguese language. Colombo: Wesleyan Mission Press.

Jackson, Kenneth David. 1990. Sing without shame: oral traditions in Indo-Portuguese verses with transcription and analysis of a nineteenth-century manuscript of Ceylon Portuguese. Amsterdam: John Benjamins and Macau: Cultural Institute of Macau.

Smith, Ian R. 1978. Sri Lanka Portuguese phonology. Trivandrum: Dravidian Linguistics Association.

———. 1979. Convergence in South Asia: a creole example. Lingua 48.

———. 1979. Substrata vs. universals in the formation of Sri Lanka Portuguese. Papers in pidgin and creole linguistics, No. 2. Canberra: Pacific Linguistics.

IAN R. SMITH

See also **Pidgins and Creoles; Tamil**

Standard Language

A standard language presupposes diglossia. Diglossia is a situation where dialectal varieties of a language in a polity function differently. The dialect adopted as the official medium of communication in formal settings becomes the standard language. Consequently, a standard language rises from a contention of several dialects. A standard language is the medium of instruction in national education, official correspondence in written communication, and mass communication in mass media like newspapers, television, films, books, and magazines.

Standard Malay in Malaysia went through interlanguage and intralanguage contentions. At the brink of independence in 1957, Malay became the national language of Malaysia and English would be phased out gradually in ten years. To equip Malay with the role of a modern language, a semigovernment body called *Dewan Bahasa dan Pustaka* was set up in 1956 to standardize the Malay spelling, and to coin Malay scientific terminology. At this interlanguage stage, Malay rose above English, Mandarin, and Tamil, among others, as the standard language in all formal settings.

Lingua franca is the common language of social or commercial communication among polyglot societies. *Johor Malay* was the lingua franca in the Malayan world. During the colonial years, *Johor Malay* was the language of publication when the press came to Malaya (preindependent Malaysia) from Singapore via Johor, a southern Malay state, to Kuala Lumpur (see Asmah). In the contention of standard language, the victory normally goes to the dialect with historical advantage. The present modern Standard Malay, however, defies the historical norm. *Johor Malay*, originating from the southern part of the Malay Peninsula, was the Standard Malay until *Kedah Malay* replaced it. Since 1988, *Kedah Malay* was decreed as Standard Malay because the writing system of this northern Malay dialect is phonemic, i.e. its spelling directly corresponds with the pronunciation. The [a] sound in an open syllable in the word-final position is pronounced as /a/ as in /baca/ (read) and /cerita/ (story) in this dialect. In comparison, /baca/ and /cerita/ are uttered as /bace/ and /cerite/ in *Johor Malay*. Due to these user- and education-friendly reasons, students and teachers were instructed to

adopt the standard [a] pronunciation in oral examinations, school lessons, and public speeches. In short, current Standard Malay in Malaysia is a result of both interlanguage and intralanguage competitions.

At the end of 1993, however, the Prime Minister of Malaysia announced that English would be the language of instruction for science and technology at the varsity level. There was an outcry over the threat of English on Standard Malay, which has been the medium of instruction from primary to tertiary education. *Internationalism* was used to describe the new era of change dawning on young polities like Malaysia, which once preoccupied itself with nationalism. Although Malay remains as Malaysia's standard language, its linguistic path transformed from the interlanguage stage to the intralanguage phase and another modern interlanguage attrition in the face of prosperity. In fact, as early as the 1980s, English-educated Malay elites were speaking *Malayish*, a mixture of Malay and English.

To give another language situation, four languages, namely English, Malay, Mandarin, and Tamil, are designated as the official languages in Singapore following a 1956 official report. Through the 1991 *Improving primary school education* report, English has eclipsed the other three official languages as the medium of instruction in Singapore. While the three official languages are taught as mother tongues in school, English dictates the academic future of the students because failing to master the English language is a disadvantage in an education system that offers English as the first language.

Interestingly, there are two varieties of English in Singapore, namely Singapore English or Singlish and Standard Singapore English. Regarded as closely connected to Chinese in terms of ethnicity and language, Singapore English lacks complicated grammatical rules and displays a fair deal of loanwords from Hokkien, a Chinese dialect, and Malay. Typical phrases of Singapore English are, '*Don't play-play*' (Do not fool around), '*Catch no ball*' (Do not understand), and the suffixation of the particle *-lah* as in '*go-lah, eat-lah, please-lah*'. In addition to providing emphasis and a sense of persuasion to the three words, the particle-*lah* can incur a negating sense of disagreement to *please*. Despite the vibrant pragmatics, i.e. situational meaning of words when in use, this variety of local English is considered better than not knowing English at all and is also thought to be associated with the lower classes and low-paid occupations. Students are required to adopt Standard Singapore English as the medium of communication and examination. Lately, as China becomes the new world economy that offers plenty of business opportunities, the go-north investment strategy gathers more momentum among Singapore entrepreneurs. Mandarin in Singapore, although has not made a significant comeback, seems to have a promising linguistic future. There is little research on the relation between Singapore and China Mandarin. Singapore exemplifies another complex inter- and intralanguage situation in the linguistic life span of standard language.

Based on the language of mass media in the United States, especially the English used in Hollywood movies and American sitcoms on television, Standard American English is the standard language that rises above African American English, Cajun English, Latino English, and Chicano English. American English is different from British English as in these contrasts: *windshield–windscreen, hood–bonnet, truck–lorry, expressway–motorway, elevator–lift, mail–post, diaper–nappy, lawyer–barrister, mean– nasty*, and *dumb–stupid*. Mistaken as American English, the word *talented* along with other American English words like *dutiable, presidential, lengthy, finalize, normalcy*, and *irregardless* were once abhorred as nonstandard English. The term Americanism, first used by John Witherspoon in 1781, refers to the use and construction of English vocabulary and sentences in the United States that are different from those in Great Britain. Americanism implies a new English dialect. When do two dialects become two different languages? By the rule of thumb, a situation of two languages arises when two dialects become mutually unintelligible. Languages like Urdu and Hindi spoken in Pakistan and India, however, are mutually intelligible; so are Danish, Norwegian, and Swedish spoken by Danes, Norwegians, and Swedes, respectively.

References

Asmah, Hj. Omar. 1992. The linguistic scenery in Malaysia. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Baugh, Albert C., and Thomas Cable. 1951, 2001. A history of the English language. London: Routledge.

Collins, James T. 1996. Devaluasi Bahasa dan Pemasaran Bahasa: Perkembangan di Negara-negara yang Berbahasa Melayu (Language devaluation and language marketing: changes in Malay-speaking countries). Jurnal Dewan Bahasa 40.

Ferguson, Charles A. 1959. Diglossia. Word 15.

Fromkin, Victoria, and Robert Rodman. 1974, 1998. An introduction to language. Fort Worth, TX: Harcourt Brace.

Gopinathan, S. 1994. Language policy changes 1979–1992: politics and pedagogy. Language, society and education in Singapore: issues and trends, ed. by S. Gopinathan, Anne Pakir, Ho Wah Kam, and Vanithamani Saravanan. Singapore: Times Academic Press.

PuruShotam, Nirmala Srirekam. 2000. Negotiating multiculturalism: disciplining difference in Singapore. Berlin: de Gruyter.

Sew, Jyh Wee. 1996. Hakisan Bahasa di Malaysia (Language erosion in Malaysia). Jurnal Dewan Bahasa 40.

Sweeney, Amin. 1987. A full hearing: orality and literacy in the Malay world. Berkeley: University of California Press.

Thomason, Sarah. 2001. Language situation. Edinburgh: University of Edinburgh Press.

JYH WEE SEW

See also: **Diglossia; Language Planning**

Story Grammar

The rediscovery of early structuralist work—for example, of Propp's research on the structure of Russian folk tales—and the development of transformational-generative grammar led to the emergence of story grammars in the 1970s. Linguists such as Morton Prince, George Lakoff, and Teun van Dijk and cognitive psychologists such as Jean Mandler and Nancy Stein contributed to this effort.

The story grammar approach holds that (1) simple stories exhibit a typical structure and (2) there is general agreement among listeners whether a given text represents a story. For example, the following set of phrases is considered a story, however trivial: 'A man was very happy, then he married a vain and domineering woman, then, as a result, he was very unhappy,' but the following phrase is not: 'Electrons are constituents of atoms.'

Among psycholinguists and cognitive psychologists, story grammar was viewed as an instance of a schema, an idealized model of the constituents of a story, and the relationship among the constituents. The story schema has been expressed in terms of rewrite rules analogous to those of transformational grammar. Representative of other story grammars, the Mandler grammar includes the following rules:

- Story → Setting and Episode(s)
- Episode → Beginning cause Complex Reaction cause Goal Path cause Ending
- Complex Reaction → Reaction cause Goal
- Goal Path → Attempt cause Outcome

This set of rules, as well as additional rules, generates a hierarchy, whose terminal elements are the surface sentences of a story.

The setting (S) describes the background that enables the events in the story. Usually, the principal characters of the story are introduced in the setting. An episode gives a story its narrative component. An episode is defined from the perspective of a protagonist who is faced with a problem and tries to solve it. The problem is triggered by events described in the beginning (B) of the episode. Then, the protagonist exhibits a reaction (R) to the problem, and he or she sets a goal (G) and makes an attempt (A) to achieve the goal. The result of the attempt is narrated in the outcome (O). The episode concludes with an ending (E), which may report the protagonists' reaction to the outcome, state a general moral, or simply emphasize

the outcome. A canonical episode includes these six constituents, where G, A, and O are obligatory in a well-formed story.

Mandler used the following dog fable to illustrate these story constituents:

- (S) It happened that a dog had got a piece of meat and was carrying it home in his mouth. Now on his way home he had to cross a plank lying across a stream.
- (B) As he crossed he looked down and saw his own shadow reflected in the water beneath.
- (R) Thinking it was another dog with another piece of meat,
- (G) he made up his mind to have that also.
- (A) So he made a snap at the shadow,
- (O) but as he opened his mouth the piece of meat fell out, dropped into the water,
- (E) and was never seen again.

The rules reflect the property of a story that the sentences do not simply represent a sequence of unrelated events. Rather, the events are causally linked as specified by the rules. A story may include multiple episodes that may be sequenced, recursively nested, or exhibit some other relation. For example, an episode may be interrupted because of an obstacle that the protagonist must overcome.

Story grammars have been criticized for various reasons. Computational linguists in the tradition of Roger Schank and Robert Abelson took the story grammar approach to task for abstracting the structure of stories from their content. Similar criticisms have also been raised vis-à-vis the syntactic approach to sentences. In each case, the principal criticism is that grammatical categories cannot be defined independent of the semantic relations of the constituents to one another. As a result, the grammars provide no source for the inferences that readers routinely make. In addition, the story schema, like other schemas, is either too general or too simple to do justice to the wide variety of narratives. Other critics argued that story grammars are technically not grammars at all because they do not specify terminal elements in the manner of transformational-generative grammar.

In defense of story grammars, it should be noted that they are not mutually exclusive from other knowledge sources at different levels of complexity, whether they are scripts, schemas, etc. And it is important to

note that the story grammar approach has produced an impressive body of empirical research validating critical assumptions about the narrative structures specified by story grammar. Consider evidence from three different paradigms: structural judgments, recall patterns, and reading times.

- Judgments of stories: Researchers had people read stories of diverse contents and divide the passages into parts. The structures produced by participants correlated well with the a priori structure of stories provided by story grammar theory.
- Recall patterns: Research revealed that people remember stories better if the stories are well formed. For other stories, people tend to generate 'improved' structures, much in the manner of subjects who 'corrected' a passage with an unfamiliar sentence structure to fit with their expectations. Further empirical studies have shown that people's recall protocols were predicted by the importance of constituents in the story hierarchy, with beginning, attempt, and outcome being best recalled.
- Reading times of story sentences in self-paced reading experiments were found to be sensitive to the episodic structure of simple two-episode

stories. Reading times adjusted for sentence length and other factors were longer at the beginning of episodes than for the remaining constituents. This pattern has been interpreted in terms of readers' effort at calling up the episode schema into memory.

The story grammar approach continues to survive in successor approaches, including causal chain theory and other theories of narrative, as described in Goldman et al.'s (1999) volume cited below.

References

Goldman, Susan R., Arthur C. Graesser, and van den Broek Paul (eds.) 1999. Narrative comprehension, casuality, and coherence: essays in honor of Tom Trabasso. Mahwah, NJ: Lawrence Erlbaum Associates.

Just, Marcel, and Patricia Carpenter. 1987. Psychology of reading and comprehension. Needham, MA: Allyn & Bacon.

Mandler, Jean. 1987. Psychological reality of story structure. Discourse Processes 10. 1–29.

Prince, Morton. 1973. A grammar of stories. The Hague: Mouton.

Propp, Vladimir. 1968. The morphology of the folktale. Bloomington: Folklore Society and Indiana University.

Van Dijk, Teun A. 1977. Text and context. London: Longman.

KARL HABERLANDT

Structural Interference

As the name signifies, structural interference denotes the intrusion of various linguistic structures of one language in the use of another. These intrusions can be at the level of phonology, morphology, syntax, discourse, pragmatics, and semantics. Some examples of structural interference in terms of transfer are given below:

1. Linguistic Transfer
1.1. Morphological: Doctor**ji**, Master**babu**, [transfer of honorific like 'ji', 'babu', from Indian languages (L1) into English (L2)]
1.2. Lexical: **/computer** eta kinib lagib /
Computer one buy should first person
I should buy a computer
[transfer of L2 lexical item, 'computer' into L1, in this case Assamese)
1.3. Word order: concerned authorities (transfer of L1 i.e. Hindi word order, i.e. adjective + noun into L2, i.e. English)

1.4. Syntactic: I am reading this book since morning (transfer of L1, i.e. Hindi syntactic feature—the use of present continuous tense for present perfect continuous—into L2, i.e. English)
1.5. Semantic: In India, use of 'auntie' for all elderly women
1.6. Discoursal: I respectfully submit the following few lines for your kind consideration (transfer of politeness rules).

2. Cultural Transfer
2.1. Lexical: What is your **good name**? (transfer of the Indian expression 'shubh-naam')
2.2. Discoursal: My nose has been cut; I can no longer show my face to anyone. (transfer of a Hindi idiomatic expression from L1 to L2).

3. Psycholinguistic Transfer

A psycholinguistic transfer is one that is determined by the learner's perception of the relationship between L1 and L2 and certain characteristics of L1 rules and items. Thus, when a German speaker is termed impolite by an English speaker, she or he perceives the two languages to be similar at the level of syntactic, discoursal, and pragmatic features. For example, when a German speaker says, 'You should close the window', in a situation where a native speaker of English will prefer using 'Can/could you close the window?', she or he is not being impolite but is transferring the politeness strategy from her or his L1 to L2 (cf. House and Kasper 1981). Again, when a Hindi speaker says, 'You are lucky, **isn't it**?', she or he uses her or his knowledge of the Hindi question tag *hein na* and transfers the rule of the application of *hein na* to English.

Structural interference is closely context bound. Increasing research evidence shows that learners can produce a significantly more fluent, grammatical, and transfer-free interlanguage (i.e. the separate linguistic system evinced when adult L2 learners attempt to express meaning in a language that they are in the process of learning) in some social contexts rather than in others.

Structural interference can also be strategic: for example, Ervin-Tripp (1969) pointed out that L2 learners might protect themselves from the consequences of sociolinguistic errors by maintaining non-native features in their speech, i.e. they might maintain structural interference to protect themselves from being assessed on the basis of native-speaker norms and expectations. Similarly, there can be deliberate structural interference to exhibit solidarity, or to serve business purposes. Researchers have noted that salespersons use different dialects and intonations depending upon their judgment of the customers' economic, linguistic, and ethnic background. In social research, some specialists suggest the use of the subjects' dialect by the researchers to make the population sample under observation feel at home. Moreover, there are different transferability criteria. The learner's assessment of the interlocutor's receptive knowledge of different language codes as well as processing constraints such as the degree of automatization are relevant factors in transfer as well as in production.

There are various theories about what, how much, why, and how a structural interference occurs. The notion of structural interference was primarily advanced by linguists interested in teaching English to learners of other languages during the 1950s. Under the influence of the structural approach in linguistics and the behaviouristic theory in psychology,

the contrastive analysts (linguists comparing the structures of two/more languages) hypothesized that language learning is habit formation and an old habit affects the formation of a new habit. Therefore, they claimed that by carrying out an explicit analysis of similar and dissimilar features of different languages, we could arrive at a foolproof theory of structural interference. When this theory was found to have loopholes, other theories emerged, stressing the processes of language acquisition. For example, the creative constructionists assume that L1 (first language/mother tongue) and L2 (second/foreign language) acquisition proceed in a similar manner. Hence, L2 acquisition is largely unaffected by L1 transfer, which in turn means structural interference is not related to language acquisition. The interlanguage hypothesis proposed by Selinker (1972), on the other hand, recognizes the major influence of the native language in the learning and use of a second language, but it also mentions some four other processes like overgeneralization of target language rules, strategies of communication, strategies of second language learning, and transfer of training. Interlanguage differs both from the native (NL/L1) and the target language (TL/L2), with fossilization (a stage where L2/TL learning stops due to pressure of communication/use) being the most significant feature. This theory of adult L2 learning is now extended to child language acquisition as well.

Research on the various aspects of structural interference has contributed significantly toward effective learning and teaching of second languages, particularly, teaching of English as a foreign language.

References

Cohen, A. 1990. Language learning. New York: Newbury House.

Ervin-Tripp, S. 1969. Sociolinguistics. Advances in experimental social psychology, Vol. 4, ed. by L. Berkowitz. New York: Academic Press.

Faerch, C., and G. Kasper (eds.) 1983. Strategies in interlanguage communication. London: Longman.

House, J., and G. Kasper. 1981. Politeness markers in English and German. Conversational routine: explorations in standardized communication situations and prepatterned speech, ed. by F. Coulmas. The Hague: Mouton Publishers.

James, C. 1980. Contrastive analysis. Harlow: Longman.

Krashen, S. 1981. Second language acquisition and second language learning. Oxford: Pergamon Press.

Parasher, S.V. 1999. Communicative styles in Indian English. Language matters, ed. by K.V. Tirumalesh. New Delhi: Allied Publishers.

———. 1977. Current trends in error analysis. JSL.

Selinker, L. 1972. Interlanguage. IRAL 10. 219–231.

SHREE VALLABH PARASHER

Structuralism

Structuralism is a mode of inquiry that consists in interpreting the phenomena it looks at as made up of relations among the various entities rather than as those entities per se. The particular units are thus defined solely by virtue of the network of relations into which they enter. They are, in other words, defined in negative terms rather than in terms of positive contents, so that any change in any one of the relationships will automatically affect the entire set of relationships within the given structure. Another way of putting this is to say that all structures constitute fully integrated systems, in which the elements are fully dependent on one another and are sensitive to the most minute of alterations taking place in any part of any given system. Because each structure is fully and exhaustively defined by the relations among the elements, it makes little sense to speak of universal structures; each structure is, as it were, a law unto itself. Yet another feature of structures in general is that they provide us with a snapshot of the phenomena rather than the evolutionary stages through which those phenomena pass; they are static and synchronic par excellence. Furthermore, many structuralists have been at pains to point out that the structures they describe are posited as such and not inherent in or latent to the phenomena themselves.

It is important to stress that structuralism did not initially emerge as a school of thought or a philosophical tendency. Yet, toward the end of the nineteenth century and in the early decades of the twentieth century, structuralism had established itself as a major force to reckon with, reaching its pinnacle of glory by the 1930s. Structuralism swept across almost all fields of inquiry, making significant contributions to the humanities and the social sciences. Among the most important names associated with the movement are Ferdinand de Saussure, Claude Lévi-Strauss, Roland Barthes, and Louis Althusser. Ferdinand de Saussure, hailed as the Father of Modern Linguistics and author of *Cours de Linguistique Générale* (1916; *A course in general linguistics*), which was published posthumously, inaugurated the structuralist revolution in linguistics. (It has, however, been argued that some of the key principles of structuralism may actually date back to thinkers such as Leibniz, Marx and Engels, Humboldt, Herder, and so on.)

It is important to point out that the term 'structuralism' came to acquire markedly different characteristics in the United States, where it is used to refer to the set of methodological principles brought to bear on linguistic analysis by the followers of Leonard Bloomfield, whose 1933 book *Language* has been referred to as the Bible of American structuralism. Among the distinguishing traits of Bloomfieldian structuralism are its stubborn aversion to questions of meaning and its close ties with behaviorist psychology. Bloomfield was also keen on making linguistics a genuinely scientific discipline, which, under the terms of the philosophy of science then in vogue, meant rigorously restricting evidence to empirically available data.

But, apart from these crucial differences, Bloomfield's structuralism was at one with its European counterpart in insisting that language be viewed as a self-contained whole and that the entities be identified relationally rather than in terms of any positive content. The key relations were those of contrast and equivalence. The sound system, or phonology, of a given language, for instance, was seen as being made up of units called 'phonemes', whose existence was predicated on the identification in the language in question of a set of contrasts with other phonemic units, each of which was to be likewise posited on the strength of the same principle. The litmus test of a contrast was a 'minimal pair', which is a pair of words identical in every respect except for the particular contrast as in *cat* and *mat*, in which the only difference is marked by the contrast between the phonemes /k/ and /m/, or in the pair *mat* and *mate*, in which the only difference consists in the change of vowels. Where two phonetically distinct items did not enter into a relationship of contrast but instead manifested what is technically known as 'complementary distribution' (that is, where one occurs, the other can never occur), the two variants were to be regarded as systematic variants of the same phoneme called 'allophones'. The same procedures were then applied to other levels of analysis, notably word structure, where the analysis yielded analogous entities such as 'morphemes' and 'allomorphs'. At the sentence level, the structuralist method yielded what is referred to as 'immediate constituent analysis'. Thus, a sentence such as *The dog barked all night* was analyzed as con-

taining the immediate constituents 'the dog' and 'barked all night'. The resulting constituents were further analyzed into the immediate constituents 'the' and 'dog' and 'barked' and 'all night', respectively, and the constituent 'all night' further on into 'all' and 'night', with the process ending once the ultimate constituents had all been identified. Finally, as what would have come as an unpleasant surprise to Bloomfield himself had he lived long enough to witness the development, the techniques of structural analysis were carried over to the analysis of meanings, paving the way for a structurally oriented theory of semantics.

Unlike its European counterpart, Bloomfieldean structuralism invested a considerable amount of effort in identifying what came to be known as 'discovery procedures', that is, a set of methodological principles with guaranteed results that would automatically and infallibly generate the right analysis from a given set of empirical data. These discovery procedures came under heavy attack from Noam Chomsky and the followers of the new paradigm of transformational-generative linguistics that he inaugurated in the 1950s. Chomsky dismissed the whole idea of discovery procedures and claimed that a linguistic theory should instead aim at attaining explanatory, rather than merely observational or descriptive, adequacy.

Despite the major differences between Bloomfieldean linguistics and generative grammar, it is nevertheless true to say that Chomsky's approach is structuralist in an extended sense of this term—although it is also the case that, thanks to the tireless criticism of the early Bloomfieldean practices by the early transformational-generative grammarians, the term 'structuralism' itself fell into disrepute and acquired pejorative connotations that survive now. Chomsky retained the basic idea of language being structured but opted for an atomistic approach to the notion of structure instead of the holistic vision that underwrote the European (mainly French) approach. The elementary building blocks of the structure were no longer negatively defined but were invested with positive attributes. Generative phonology adopted as its basic building blocks 'distinctive features' that were universal and not language specific. The so-called phonemes were from now on seen as merely contingent and language-specific combinations of these features. This important theoretical intervention permitted Chomsky and his followers to posit a universal base for all languages, shifting the focus of attention to language as an innate species-specific attribute of man and redefining linguistics itself as a branch of cognitive psychology.

It should not be concluded from the foregoing remarks that structuralism in its classic sense is by now a spent force. Quite on the contrary, Saussure's teachings continue to inspire generations of scholars and is still a point of reference for so-called post-structuralism, whose advocates, far from rejecting *tout court* the legacy of Saussure's teachings, use them as a springboard for further advancing his insights, albeit to logical consequences unimagined by the Swiss savant. As for Bloomfield's legacy, it too arose from the ashes and gained a further lease on life in the form of what is referred to as neo-Bloomfieldeanism, notably in the work of Bernard Bloch (1907–1965).

References

Abraham, S., and F. Kiefer. 1966. A theory of structural semantics. The Hague: Mouton.

Bartsch, R., and T. Vennemann (ed.) 1973. Linguistics and neighboring disciplines. Amsterdam: North-Holland.

Bloomfield, Leonard. 1933. Language. New York: Holt, Rinehart & Winston.

Cassirer, E. 1945. Structuralism in modern linguistics. Word 1.

Chafe, Wallace L. 1971. Meaning and structure of language. Chicago: University of Chicago Press.

Chomsky, Noam A. 1957. Syntactic structures. The Hague: Mouton.

Culler, Jonathan. 1975. Structuralist poetics: structuralism, linguistics, and the study of poetics. London: Routledge & Kegan Paul.

Derrida, Jacques. Of grammatology, English translation by G.C. Spivak. Baltimore, MD: Johns Hopkins University Press.

Dinneen, Francis. P. 1967. An introduction to general linguistics. New York: Holt, Rinehart & Winston, Inc.

Francis, W. Nelson. The structure of American English. New York: Harcourt, Brace & World.

Hall Jr., Robert A. (ed.) 1987. Leonard Bloomfield: essays on his life and work. Amsterdam and Philadelphia: John Benjamins.

Hockett, Charles. F. A course in modern linguistics. New York: Macmillan.

Hymes, D., and Fought, J. American structuralism. The Hague: Mouton.

Joseph, J.E. Bloomfield's Saussureanism. Cahiers Ferdinand de Saussure, 43.

Lévi-Strauss, Claude. 1963. Structural anthropology. New York: Basic Books.

Saussure, Fedinand de. 1916. Cours de Linguistique Générale. Paris: Payot, English translation by Wade Baskin, A course in general linguistics. New York: Philosophical Library, 1959.

Wells, R. 1957. De Saussure's system of linguistics. Readings in linguistics, ed. by M. Joos. Washington, DC: American Council of Learned Societies.

KANAVILLIL RAJAGOPALAN

See also **Bloomfield, Leonard; Chomsky, Noam; Lévi-Strauss, Claude; Saussure, Ferdinand de**

Stylistics

Stylistics is the description and analysis of the variability of linguistic forms in actual language use. The concepts of 'style' and 'stylistic variation' in language rest on the general assumption that within the language system, the same content can be encoded in more than one linguistic form. Operating at all linguistic levels (e.g. lexicology, syntax, text linguistics, and intonation), stylisticians analyze both the style of specific texts and stylistic variation across texts. These texts can be literary or nonliterary in nature. Generally speaking, style may be regarded as a choice of linguistic means; as deviation from a norm; as recurrence of linguistic forms; and as comparison.

Considering style as choice, there are a multitude of stylistic factors that lead the language user to prefer certain linguistic forms to others. These factors can be grouped into two categories: user-bound factors and factors referring to the situation where the language is being used. User-bound factors include, among others, the speaker's or writer's age; gender; idiosyncratic preferences; and regional and social background. Situation-bound stylistic factors depend on the given communication situation, such as medium (spoken vs. written); participation in discourse (monologue vs. dialogue); attitude (level of formality); and field of discourse (e.g. technical vs. nontechnical fields). With the caveat that such stylistic factors work simultaneously and influence each other, the effect of one, and only one, stylistic factor on language use provides a hypothetical one-dimensional variety. Drawing on this methodological abstraction, stylistic research has identified many correlations between specific stylistic factors and language use. For example, noun phrases tend to be more complex in written than in spoken language in many speech communities, and passive voice occurs much more frequently in technical fields of discourse than in nontechnical ones.

Style, as deviation from a norm, is a concept that is used traditionally in literary stylistics, regarding literary language as more deviant than nonliterary language use. This not only pertains to formal structures such as metrics and rhyme in poems but to unusual linguistic preferences in general, which an author's poetic license allows. Dylan Thomas's poetry, for example, is characterized by word combinations that are semantically incompatible at first sight and, thus, clearly deviate from what is perceived as normal (e.g. *a grief ago, once below a time*). What actually constitutes the 'norm' is not always explicit in literary stylistics, since this would presuppose the analysis of a large collection of nonliterary texts. However, in the case of authorship identification, statistical approaches were pursued at a relatively early stage. For example, by counting specific lexical features in the political letters written by an anonymous Junius in the 1770s and comparing them with a large collection of texts from the same period, and with samples taken from other possible contemporary authors, the Swedish linguist Ellegård could identify, in the 1960s, the most likely author of those letters.

The concept of style as recurrence of linguistic forms is closely related to a probabilistic and statistical understanding of style, which implicitly underlies the deviation-from-a-norm perspective. It had already been suggested in the 1960s that by focusing on actual language use, stylisticians cannot help describing only characteristic tendencies that are based on implicit norms and undefined statistical experience in, say, given situations and genres. In the last resort, stylistic features remain flexible and do not follow rigid rules, since style is not a matter of grammaticality, but rather of appropriateness. What is appropriate in a given context can be deduced from the frequency of linguistic devices in this specific context. As for the analysis of frequencies, corpus linguistic methods are becoming increasingly important. With the advent of personal computers, huge storage capacities, and relevant software, it is now possible to compile very large collections of texts (corpus (singular), corpora (plural)), which represent a sample of language use in general, and thus enable exhaustive searches for all kinds of linguistic patterns within seconds. This methodology is based on the general approach of style as probability, by allowing for large-scale statistical analyses of text. For example, by using corpora, the notion of text-type—defined by co-occurrences of specific linguistic features—has been introduced to complement the extralinguistic concept of 'genre'. The linguistically defined text types contradict traditionally and nonempirically established genre distinctions to a considerable extent. In particular, many spoken and written genres resemble each other linguistically to a far greater extent in terms of text-types than previously assumed.

Style as comparison puts into perspective a central aspect of the previous approaches. That is, stylistic

analysis always requires an implicit or explicit comparison of linguistic features between specific texts, or between a collection of texts and a given norm. In principle, stylistically relevant features such as style markers may convey either a local stylistic effect (e.g. an isolated technical term in everyday communication) or, in the case of recurrence or co-occurrence, a global stylistic pattern (e.g. specialized vocabulary and passive voice in scientific texts).

From the multitude of linguistic approaches to style, two linguistic schools of the twentieth century have exerted the most decisive influence on the development, terminology, and the state of the art of stylistics: the Prague School and British Contextualism.

The central dictum of Prague School linguistics, going back to the Bauhaus School of architecture, is 'form follows function'. Firmly established since the 1920s, some of this dictum's most important proponents are Lubomír Doležel, Bohuslav Havránek, Roman Jakobson, and Jan Mukařovský. These linguists have paid particular attention to situation-bound stylistic variation. A standard language is supposed to have a communicative and an esthetic function that result in two different 'functional dialects': prosaic language and poetic language. More specific functional dialects may, of course, be identified; for example, the scientific dialect as a subclass of prosaic language, which is characterized by what is called the 'intellectualization of language'—lexicon, syntax, and reference conform to the overall communicative function that requires exact and abstract statements.

A very important notion is the distinction between 'automatization' and 'foregrounding' in language. Automatization refers to the common use of linguistic devices which does not attract particular attention by the language decoder, for example, the use of discourse markers (e.g. *well*, *you know*, *sort of*, *kind of*) in spontaneous spoken conversations. Automatization thus correlates with the usual background pattern, or the norm, in language use—it encompasses those forms and structures that competent language users expect to be used in a given context of situation. Foregrounded linguistic devices, on the other hand, are usually not expected to be used in a specific context and are thus considered conspicuous—they catch the language decoder's attention (e.g. the use of old-fashioned and/or very formal words such as *epicure*, *improvident*, and *whither* in spontaneous spoken conversations). Foregrounding thus captures deviations from the norm. It is obvious that what is considered as automatized and foregrounded language use depends on the communication situation at hand. In technical fields of discourse, for instance, specialized vocabulary items tend to be automatized (e.g. *lambda marker* in molecular biology), but in everyday communication become foregrounded devices.

A different, although conceptually similar, tradition of linguistic stylistics was established by British linguists in the 1930s and came to be called British Contextualism. The most important proponents of British Contextualism include John Rupert Firth, M.A.K. Halliday, and John Sinclair. Their work is characterized by a clear focus, firstly, on the social context in which language is used and, secondly, on the indepth observation of natural language use. From the point of view of British Contextualists, linguists need to describe authentic language use in context and should not confine themselves to invented and isolated sentences. Additionally, linguistics is not considered as an intuition-based study of abstract systems of form as, for example, in the merely formal description of autonomous syntactic rules (as in Chomsky's approach to language), but as the observation-based and empirical analysis of meaning encoded by form. This approach allows for insights into the immense variation within language. It is a fact that depending on the context of situation, all speakers use different 'registers' (i.e. different styles of language, depending on the topic, the addressee, and the medium in a given context of use). Note that there is, of course, a clear correspondence between the concept of register and the Prague School's notion of functional dialect. Although largely abandoned by mainstream linguists in the 1960s and 1970s due to the prevailing Chomskyan school of thought, it had already been suggested by Firth in the 1950s that large collections of text were a prerequisite for an empirical approach to stylistic variation. Thus, it does not come as a tremendous surprise that, among others, Sinclair set out to develop computerized corpora that could be used as empirical databases.

With corpus linguistics now a standard methodology, stylistic analyses have reached an unprecedented degree of explanatory adequacy and empirical accuracy. For example, stylistic features that are beyond most linguists' scope of intuition, such as the nonstandard use of question tags in English-speaking teenagers' talk, are now feasible in quantitative terms. More importantly, there is no longer a bias toward foregrounded phenomena that tend to catch the linguist's attention. A computer, in contrast, does not distinguish between conspicuous and common phenomena and provides an exhaustive array of all kinds of patterns, depending solely on the search query. Thus, the fuzzy concept of 'norm' is about to be put on an empirical footing since the accessible corpus norm represents the norm of a language as a whole.

Stylistics is a linguistic branch that is immediately relevant to foreign language teaching. This applies to both linguistic and literary stylistics. Language learners must know which linguistic devices are preferred by native speakers in specific contexts. Without such a

linguostylistic competence, communication errors may be made in interacting with native speakers, such as using highly formal words in informal settings. Also, learners must have command of text-typological knowledge, which is important, for example, in writing essays. As for literary texts, language learners should acquire a firm understanding of those levels of description where stylistic variation may occur (e.g. by analyzing Hemingway's syntactic simplicity and, moreover, its function).

It should be noted that a specific style is sometimes ascribed to a language in its entirety. Although the underlying norms remain largely unspecified, general tendencies of stylistic preference differ across languages. This is particularly important for translators, but also for language learners. It is, for instance, common for German students of English to transfer the German style of academic writing, which is characterized by heavy noun phrases, to their English essays.

As with any other linguistic branch, stylistics is very much a work in progress. This is because the object of inquiry constantly grows, evolving new and specialized fields of discourse (e.g. genetic engineering, computer sciences). Furthermore, new aspects of stylistic variation come into existence, such as e-mails, a now widely used genre that seems to blur the traditional distinction between spoken and written language. As for empirical approaches to style, new corpora make it possible to address questions of style not possible before. Also, recent theoretical developments will no doubt widen the scope of stylistics. Drawing on British contextualists' distinction between language substance (that is, sound waves in the phonic medium and printed paper in the graphic medium) and language form (that is, anything that can be transferred from one medium into the other), it has been suggested that stylistic analyses should clearly distinguish between medium-dependent and medium-independent stylistic variation. Intonation, for example, is bound to the phonic medium and shows stylistic variation that cannot be mapped onto punctuation in a straightforward and monocausal way. With regard to the graphic substance, English orthography, albeit highly standardized, is also affected by stylistic variation, as deliberate misspellings in the language of advertising and popular culture (e.g. *2* for *to/two/too*, *lynx* for *links*) reveal. On the other hand, words and syntax are linguistic devices that, in principle, are subject to transfer between media, although there are clear medium-dependent preferences of lexical and syntactic choice that need to be investigated further.

The objective and unbiased approach to stylistic variation in authentic language use is a cornerstone of modern descriptive linguistics. Unlike traditional grammar, it clearly rejects the normative prescription of one specific style.

References

Biber, Douglas. 1989. Variation across speech and writing. Cambridge: Cambridge University Press.

Coupland, Nikolas (ed.) 1988. Styles of discourse. London: Croom Helm.

Enkvist, Nils Erik. 1973. Linguistic stylistics. The Hague: Mouton.

Esser, Jürgen. 1993. English linguistic stylistics. Tübingen: Niemeyer.

———. 2000. Medium-transferability and presentation structure in speech and writing. Journal of Pragmatics 32.

Garvin, Paul L. (ed.) 1964. A Prague school reader on esthetics, literary structure and style. Washington: Georgetown University Press.

Halliday, M.A.K. 1978. Language as a social semiotic: the social interpretation of language and meaning. London: Arnold.

Joos, Martin. 1961. The five clocks: a linguistic excursion into the five styles of English usage. New York: Harcourt, Brace and World.

Leech, Geoffrey, and Michael Short. 1981. Style in fiction: a linguistic introduction to English fictional prose. London: Longman.

Semino, Elena, and Jonathan Culpeper (eds.) 2002. Cognitive stylistics: language and cognition in text analysis. Amsterdam: Benjamins.

Sinclair, John. 1991. Corpus, concordance, collocation. Oxford: Oxford University Press.

Weber, Jean Jacques (ed.) 1996. The stylistics reader: from Roman Jakobson to the present. London: Arnold.

JOYBRATO MUKHERJEE

See also **Firth, John Rupert; Halliday, Michael Alexander Kirkwood**

Sumerian

Sumerian was a language spoken in the south of Ancient Mesopotamia (modern Iraq) and is most likely first attested in the archaic texts from Uruk and Jemdet Nasr (from the end of the fourth millennium to the middle of the third). By the end of the third millennium, Sumerian had died out for the most part as a spoken

language. However, it was still used in a wide variety of literary, scholarly, and religious genres, and was preserved in writing until the practical disappearance of the Mesopotamian civilization. Sumerian is an ISOLATE, i.e. it is not related to any other language or language family. Thus, our knowledge of Sumerian grammar and lexicon is mostly based on a large number of bilingual texts (in Sumerian and Akkadian), as well as a stream of scribal and scholastic traditions materialized in a corpus of lexical lists and grammatical texts.

Phonology and the Writing Interface

Mesopotamian CUNEIFORM is the logosyllabic script used for Sumerian. For instance, the sign AN (originally a star) could write the word **an** 'sky, heaven', as well as **dingir** 'god'. As a consequence of this and other processes, cuneiform signs are polyphonic, i.e. they have several readings (e.g. the sign NE can be read {ne}, {de$_3$}, {bi$_2$}, etc.). On the other hand, the writing interface shows HOMOPHONY, because different signs can have the same reading, such as {e}, {e$_2$}, {e$_3$}, {e$_4$}..., {e$_{11}$} (numerical indexes are used to distinguish between homophonous signs). Furthermore, scribes used determinatives preceding or following some nouns and names, simply as reading aids: KI 'place' in **lagaš^{ki}**; DINGIR 'deity' in **^{d}nanna**, etc.

Our knowledge of Sumerian PHONOLOGY is limited by the nature of the writing interface. The vocalic system is quite simple: /a/, /i/, /e/, /u/. The possible existence of an /o/ vowel remains unsubstantiated. Vocalic length does not seem to be phonological. The consonantal system includes three series of stops (/b/ and /p/; /d/ and /t/; and /g/ and /k/), three sibilants (/s/, /z/, and /š/ = IPA ʃ), two laterals (/l/ and /r/), three nasals (/m/, /n/, and /g̃/ or /ŋ/), and a uvular or pharyngeal of undefinable nature (/h/ or /ḫ/). Furthermore, the existence of different 'extra phonemes' (segments that would not be immediately obvious in the writing interface) has been proposed. Among these, the velar nasal /g̃/ or /ŋ/ is now commonly accepted by all Sumerologists—for instance, both {mu} and {gu$_{10}$} are written with the same sign, MU—as well as /d^r/ (most likely /ř/). To further complicate things, some final consonants (*Auslaut*) seem to drop (**kala** 'mighty' < /kalag/), but the same signs have readings with and without *Auslaut*: **kala** = **kalag**; **dug$_4$** = **du$_{11}$** 'to speak'.

Our knowledge of Sumerian PHONOTACTICS is limited. Although the cuneiform system, as most syllabaries, precludes the writing of initial and final consonantal clusters, it is quite likely that Sumerian did have them. In compounds, medial consonantal clusters were also sometimes difficult to indicate: **ki-**

sikil /kiskil/ ('girl'), **gir$_5$-gir$_5$-(r)e** /gigre/ ('to sink, to dive'), etc. Some have suggested that Sumerian may have had lexical tones as a suprasegmental feature, which allegedly would explain the sometimes high number of homophonic terms. However, other factors can explain this: different *Auslaut* consonants, consonantal clusters, etc.

VOWEL HARMONY occurs with several verbal prefixes, and traces of it can be found in some nouns. The explicit writing of vowel harmony developed through time in diverse ways. For instance, in the case of the verbal modal prefix /he-/ or /ha-/ (see below), the orthographic evolution went from one allomorph (**he$_2$-**), to two (**he$_2$-** and **ha-**), and eventually to three (**he$_2$-**, **ha-**, and **hu-**).

Nominal Morphology

Sumerian is an AGGLUTINATIVE language. The grammatical gender is based on an opposition between animate and inanimate nouns, but has no specific morphological marker and only surfaces in morphosyntactical relations between pronouns and their antecedents. Grammatical number (plural vs. singular) does not need to be marked in writing (**lugal** 'king' or 'kings'), but can be made explicit through suffixation (**lugal-e-ne** /lugal-ene/ 'kings') or reduplication (**lugal-lugal** 'kings'). The absence of marker and lexematic reduplication are probably simple orthographic conventions to write the plural, as may be indicated by the construction of plurals with reduplication of the adjective (**dingir gal-gal** = *god-great-great* 'great gods') or with reduplication and an additional suffix (**dingir gal-gal-e-ne** = *god-great-great*-PLURAL 'great gods').

In the nominal system, there are ten cases, which are marked by attaching suffixes to Noun Phrases (NPs). Moreover, cases can also be indicated by inserting prefixes in verbal forms (all elements preceding the stem are called prefixes):

	Nominal suffixes	*Prefixes in verbal chain*
ERGATIVE	/-e/	
ABSOLUTIVE	/-∅/	
GENITIVE	/-ak/	
DATIVE	/-ra/	/-na-/
LOCATIVE	/-a/	/-ni-/
COMITATIVE	/-da/	/-da-/
TERMINATIVE	/-eše/ ({-še$_3$})	/-ši-/
ABLATIVE–INSTRUMENTAL	/-ta/	/-ta-/ or /-ra-/
LOCATIVE–TERMINATIVE	/-e/	/-e-/ (or /-i-/)
EQUATIVE	/-gin/ ({-gin$_7$} = GIM)	

NPs are usually called 'nominal chains' by Sumerologists, because all the suffixes are heaped at the very end of the last syntagm:

dumu lugal kalam-ma-ka-ke₄-ne-ra
/dumu lugal kalam-ak-ak-ene-ra/
son-king-nation-GEN-GEN-PL-DAT
'for the son of the kings of the nation (=Sumer)'

Nonetheless, the syntagmatic structure is not so much that of a chain as it resembles a series of Chinese boxes (Zólyomi 1996):

ᵦ[dumu ₐ[lugal kalam-ak]ₐ-ak-ene]ᵦ-ra

Sumerian is an ERGATIVE language: the subject of an intransitive verb presents the same marker as the object of a transitive verb (the absolutive case), while the subject of a transitive verb presents a marker (the ergative case) that is different from that of the intransitive verb. Thus, Sumerian has /-e/ as the ergative suffix, and /-∅/ as the marker of absolutive case:

lugal-e e₂ mu-un-du₃	'the king built the temple'
lugal i₃-tuš	'the king sat down'
nin-e in-tud-en	'the queen bore me'
nin i₃-tuš	'the queen sat down'

In fact, Sumerian exhibits split ergativity in its morphology. The ergative alignment is strictly followed only in the nominal system. Independent personal pronouns, imperatives, cohortative verbal forms, and a few nonfinite verbal constructions exhibit an accusative alignment. The system of verbal agreement shows a similar split: the *ḫamṭu* forms (perfective) are ergative, whereas the *marû* forms (imperfective) show an accusative pattern.

The Verbal System

Verbal stems are usually divided into two major categories: *ḫamṭu* ('quick, sudden' → perfective) and *marû* ('slow, fat' → imperfective). However, these two labels, rather than native Sumerian grammatical categories, reflect the understanding of the Sumerian verb by Akkadian-speaking scribes. In an early grammatical text, **lugud** ('short') occurs instead of *ḫamṭu*, and **gid₂** ('long') instead of *marû* (Civil 2002)—the same two terms were used by ancient scribes in naming signs according to their shapes. Thus, the same labels, when used to designate verbal stems, refer to the shape of the stems, which would be regarded as long because of reduplication and other possible changes. These two stems marked tense and aspect, perfective and imperfective, as well as mood (deontic versus epistemic) when occurring with certain modal prefixes (see below).

The problem of the marking of the so-called *marû* stem is still widely discussed. It is likely that all verbs had two stems. Affixation verbs perhaps marked the *marû* stem with an affix /-e/; reduplication verbs (like **gar** 'to place') marked it with partial reduplication (**ga₂-ga₂**), as opposed to complete reduplication in *ḫamṭu* forms (**gar-gar**); alternating verbs (such as **e₃** 'to go out') with their 'expanded form' (/e₃-d/); and suppletive or complementary verbs (such as **dug₄**) with completely different lexemes (**e**) (see Yoshikawa 1993: 1–56, 95–104, 114–26). However, it is also possible that many verbs did not have two different stems, and that the only way to distinguish *ḫamṭu* from *marû* in those verbs was through concord (i.e. through pronominal affixes). According to this, /-e/ would not mark *marû* but agreement for the 3rd sg. subjects of transitive *marû* forms (intransitive *marû* forms have /-∅/ in the 3rd sg.).

Verbal forms can present both prefixes and suffixes. Although almost no actual verbal form exhibits all the possible affixes, an ideal table of slots would be as in Table 1.

Whereas some MODAL PREFIXES always indicate the same mood (**ga-** → cohortative; **nu-** →negative; **ša-** → affirmative), others can mark either deontic or epistemic modality, depending on their interaction with the other elements within the TAM system (Civil 2000):

/ha-/ + *ḫamṭu*	→ affirmative	**ha-na-sum** 'I have indeed given'
/ha-/ + *marû*	→ precative	**hu-mu-hul₂-e** 'may you rejoice'
/na-/ + *ḫamṭu*	→ affirmative	**nam-mi-gub** 'he set indeed'
/na-/ + *marû*	→ prohibitive	**na-ab-pad₃-de₃** 'do not tell'
/bara-/ + *ḫamṭu*	→ negative affirmative	**ba-ra-ra-dug₄** 'I have never said to you'
/bara-/ + *marû*	→ vetitive	**ba-ra-pad-re₆** 'he shall not destroy'

Other prefixes in this slot do not really indicate modality: **u₃-** marks anteriority and /inga-/ is actually a proclitic connective particle.

The number, function, rank, compatibility, and morphophonemic shape of the so-called CONJUGATION PREFIXES are still a matter of discussion. For instance, /-m-/ is probably the noninitial allomorph of /mu-/, which explains why both prefixes cannot occur together. It is possible to argue for a rather simple system of conjugation prefixes with only four morphemes: /ba-/;

TABLE 1 Structure of the Sumerian Verbal Chain

1	2	3	4	5	6	7	8	9	10
modal prefixes	connective prefix	"conjugation" prefixes	dative	dimensional infixes	pronominal prefixes	STEM		pronominal suffixes	nominalizer
∅-		i- (V-)							
he₂-		mu- (m-)	-a- (1 sg)	-ši- TERMINATIVE	-∅-/-e- (2 sg)			-en / -en	
ga-	inga-	ba- (b-)	-ra- (2 sg)	-ni- LOCATIVE	-n- (3 sg an)			-en / -en	
bara-		im-ma-	-na- (3 sg)	-da- COMITATIVE	-b- (3 sg in)		-ed	-∅ / -e	-a
na-		a-	-me- (1 pl)	-ta- ABLAT.-INST.				-enden / -enden	
ša-			[*-re- (2 pl)]	-e- / -i- LOCAT.-TERM.				-enzen / -enzen	
u₃-			-ne- (3 pl)					-eš / -ene	
nu-									
al-									

/imma-/; /i-/ or /V-/; and /mu-/. The prefix /bi-/ would be a combination of the prefix /ba-/ followed by the locative-terminative infix, and /imma-/ would be a reduplication of /mu-/ (Karahashi 2000).

All verbal forms seem to start with an obligatory conjugation prefix (/mu-/, /ba-/, or /i-/). The choice of prefix seems governed by FOCUS (Vanstiphout 1985). The prefixes /mu-/ and /ba-/ are mutually exclusive and complementary: /mu-/ is focused for a person but not for a locus, while /ba-/ is focused for a locus but not for a person; and /i-/ is not focused; and hence it is indifferent to this opposition. In discourse, the /i-/ prefix is preferred for supportive, nonsubstantial material (background), but the foreground of regular narrative

a verbal prefix. Moreover, in order to agree with plurals, the pronominal prefixes can occur together with plural pronominal suffixes.

The PRONOMINAL SUFFIXES are identical for all verbal forms with the exception of the 3rd person singular and plural. A 3rd person in the absolutive case shows concord with /-∅/ in the singular and with /-eš/ in the plural. An ergative with a *marû* form shows concord with /-e/ in the singular and /-ene/ in the plural. The absolutive case of a transitive *marû* construction (i.e. the accusative case with *marû*) does not agree with the pronominal suffixes but with the pronominal prefixes (/-b-/ and /-n-/). The distribution of pronominal elements in verbal forms is as follows:

	transitive ḫamṭu	intransitive ḫamṭu	transitive marû	intransitive marû
subject	prefix (+ suffix in pl.)	suffix (3rd -∅/-eš)	suffix (3rd -e/-ene)	suffix (3rd -∅/-eš)
object	suffix (3rd -∅/-eš)		prefix	

discourse is marked by /mu-/ or /ba-/ according to focus:

+ FOCUS			- FOCUS
/mu-/	+ PERSON	- LOCUS	/i-/
/ba-/	- PERSON	+ LOCUS	

The conjugation prefix /a-/ seemingly occurs in nonagentive passive constructions (**an-na-sum** 'it was given to him'). This prefix defocalizes the agent of the sentence, shifting the focus to any other argument (dative, comitative, ablative, and so forth); see Yoshikawa (1995). The occurrence of /a-/, however, may also correspond to local or diachronic dialects (or even to scribal idiolects), since in some texts it is extremely rare or completely unattested, while in others it is quite frequent.

The DIMENSIONAL INFIXES mark case functional relations between the verb and NPs that may be explicit or merely implicit in the sentence. The PRONOMINAL PREFIXES normally agree with the subject of transitive ḫamṭu forms (ergative) and the subject of both transitive and intransitive *marû* forms—the latter showing an accusative alignment. Furthermore, they can also specify that a dimensional prefix (terminative, comitative, etc.) refers to a 2nd or 3rd person, as well as the gender (/-n-/ for animate and /-b-/ for inanimate) of the verbal object or any syntactical argument marked with

With regard to the choice of prefixes vs. suffixes, the ḫamṭu forms follow an ergative pattern, whereas the *marû* ones show accusativity. Nonetheless, the pronominal suffixes used with the *marû* forms do point to an opposition between ergative subject and absolutive subject for the 3rd singular and plural (i.e. they ultimately have an ergative alignment).

The SUFFIX /-ED/ can immediately follow the verbal stem and precede the pronominal suffix. Some consider the /-e-/ in /-ed/ the marker of *marû* (see above). This suffix is much more frequent in nonfinite than in finite verbal constructions (in which it indicates future in diverse modalities): **e₂-mu lu₂ i₃-bur₃-de₃** 'someone could break (/i-bur-(e)d-e/) into my house'; **e₂ du₃-de₃ igi-zu u₃ dug₃-ga nu-ši-ku₄-ku₄** 'in order to build (/du₃-(e)d-e/) the house you will not let sweet sleep enter your eyes'; and **ur-sag e₂-a-na ku₄-ku₄-da-ni ud me₃-še₃ gu₃ ga₂-gar-am₃** 'at entering (/ku₄-ku₄-(e)d-a-ani/ = *enter-enter*-ED-NOMINALIZER-*his*) his house, the warrior was a storm roaring towards battle'. The suffix /-ed/ is never written as such: /-e-/ is written almost exclusively after a consonant, and /d/ is written only when followed by a vowel.

The IMPERATIVE exhibits a reverse order of verbal constituents: it begins with the stem, which is followed by all the prefixes; e.g. **sum-ma-ab** /sum-mu-a-b/ 'give (sg.) it to me', **sum-ma-ab-ze₂-en** /sum-mu-a-b-enzen/ 'give (pl.) it to me'. This phenomenon is similar to the switch from proclisis to enclisis in the imperative in other languages; e.g. Spanish *me lo das* ('you give it to me') vs. *dámelo* ('give it to me').

Morphosyntax

The NOMINALIZER suffix /-a/ can be attached to both nonfinite and finite verbal forms, and can be followed by case endings and pronouns. When the nominalized verbal form agrees with an explicit (such as **lu₂** 'man'→ who/that) or implicit noun that has an antecedent in another sentence, it constitutes the equivalent to an English relative clause: **ensi₂ lu₂ e₂-ninnu in-du₃-a e₂-uru-gir₂-su^ki-ka-ni mu-na-du₃** 'the *ensi*-ruler that built the Eninnu (**lu₂... in-du₃-a** ← /i-du₃-a/), built her temple of the city of Girsu'. Due to the high number of grammatical functions explicitly marked, there is no obvious syntactical opposition between parataxis and hypotaxis (the latter would correspond entirely to nominalized verbal forms). For instance, the word order tends to be almost always Subject–Object–Verb in all sentences.

The verb TO BE (**me**) has a basic essential meaning and does not indicate existence, for which **gal₂** ('to be there, to exist') is used: **pi-lu₅-da ud-bi-ta e-me-a** 'these were (/i-me-a(m)/) the conventions of old times'. It occurs more frequently in the form of an enclitic copula, with the following suffixes:

1st sg.	**-me-en**	1st pl.	**-me-enden**
2nd sg.	**-me-en**	2nd pl.	**-me-enzen**
3rd sg.	**-(a)-m**	3rd pl.	**-me-eš**

E.g., **digir-ra-ni ^dšul-utul₁₂-am₆** 'his god is Šulutul', **ama-mu ze₂-me** 'you are (/ze₂-me-e(n)/) my mother.'

The PRONOMINAL SYSTEM follows an accusative alignment. The pronominal subjects of both transitive and intransitive verbs present the same marker, /-e/:

	1st sg.	2nd sg.	3rd sg.	3rd pl.
SUBJECT	**ga₂-e (me-e)**	**za-e (ze)**	**e-ne**	**e-ne-ne**
DATIVE	**ga₂-a-ra (ma-a-ra)**	**za-a-ra**	**e-ne-ra**	**e-ne-ne-ra**
TERMINATIVE	**ga₂(-a/e)-še₃**	**za(-a/e)-še₃**	**e-ne-še₃**	**e-ne-ne-še₃**
COMITATIVE	**ga₂(-a/e)-da**	**za(-a/e)-da**	**e-ne-da**	**e-ne-ne-da**
EQUATIVE	**ga₂(-a/e)-gin₇**	**za(-a/e)-gin₇**	**e-ne-gin₇**	**e-ne-ne-gin₇**

The POSSESSIVE SUFFIXES can be attached to NPs as well as to nominalized verbal forms:

1st sg.	**-mu (-gu₁₀)** 'my'	1st pl.	**-me** 'our'
2nd sg.	**-zu** 'your'	2nd pl.	**-zu-ne-ne**, **-zu-e-ne-ne**, **-zu-ne** 'your'
3rd sg. an.	**-a-ni** 'his, her'	3rd pl.	**-a-ne-ne** 'their'
3rd sg. in.	**-bi** 'its'		**-bi** 'their' (probably collective)

Sumerian has a number of COMPOUND VERBS, i.e. combinations of a verb and a direct object that become a syntactic and semantic unit: : **igi — bar** ('to look at' ← **bar** 'to open' + **igi** 'eye'), **ki — ag₂** ('to love' ← **ag₂** 'to measure' + **ki** 'place'), etc. The second object of a compound verb is very frequently in the locative–terminative case: **nig₂-dug₃-ge al na-an-ga-am₃-mi-in-dug₄** 'sweet things (/nig₂-dug₃-e/) she has indeed also wished (/na-inga-bi-n-dug₄/, **al — dug₄** 'to desire, wish')'. There are some double compound verbs, in which a whole compound verb becomes the nominal part of a compound verb, whose verbal member is an 'auxiliary' verb, such as **ak** ('to do') and **dug₄** ('to say'): **šu tag — dug₄** ← **šu — tag** 'to cover, decorate' (← **šu** 'hand' + **tag** 'to touch').

A Sumerian Genderlect or Genrelect?

Sumerian is called **eme-gir₁₅** (perhaps 'native tongue') in native Sumerian sources. In some Mesopotamian scholarly texts, a few lexical items and grammatical forms are identified as **eme-sal** (perhaps 'fine language'). It has been argued that **eme-sal** was a women's language (*Frauensprache*) or genderlect, especially because the sign SAL can also be read **munus** 'woman'. The fact is that **eme-sal** is attested in compositions of very specific genres: cultic songs performed by **gala**-priests; diverse texts containing Inanna's speech; some laments over the destruction of cities; a lullaby; about 30 proverbs or short sayings from rhetorical collections; an unpublished composition ('The song of the millstone'); and the 'Dialogues between two women'. No text is entirely written in **eme-sal**, and there is no true consistency in its use; hence, an otherwise 'main-dialect' text may present some scattered **eme-sal** words. In most cases, the occurrence of **eme-sal** forms seems determined mostly by the genre of the text.

References

Attinger, Pascal. 1993. Eléments de linguistique sumérienne: La construction de **du₁₁/e/di** «dire». Fribourg/Göttingen: Editions Universitaires/Vandenhoeck&Ruprecht.

Civil, Miguel. 2000. Modal prefixes. Acta Sumerologica 22, in press.

———.2002. The forerunners of *marû* and *ḫamṭu* in Old Babylonian. Riches hidden in secret places: ancient Near Eastern studies in memory of Thorkild Jacobsen, ed. by Tzvi Abusch, 63–71. Winona Lake, IN: Eisenbrauns.

Edzard, Dietz Otto. 2003. Sumerian grammar. Leiden: Brill.

Karahashi, Fumi. 2000. The locative-terminative verbal infix in Sumerian. Acta Sumerologica 22, in press.

Michałowski, Piotr. 1980. Sumerian as an ergative language. Journal of Cuneiform Studies 32. 86–103.

———.2004. Sumerian. The Cambridge encyclopedia of ancient languages, ed. by Roger D. Woodard, 19–59. Cambridge: Cambridge University Press.

Thomsen, Marie-Louise. 1984. The Sumerian language. Copenhagen: Academic Press.

Vanstiphout, Herman L.J. 1985. On the verbal prefix /i/ in standard Sumerian. Revue d'Assyriologie 79. 1–15.

Yoshikawa, Mamoru. 1993. Studies in the Sumerian verbal system. Hiroshima: The Middle Eastern Culture Center in Japan.

———.1995. The Sumerian verbal prefix /a-/. Acta Sumerologica 17. 299–307.

Zólyomi, Gábor. 1996. Genitive constructions in Sumerian. Journal of Cuneiform Studies 48. 31–47.

GONZALO RUBIO

Swadesh, Morris

Like his teacher Edward Sapir, Morris Swadesh was a prolific data-gatherer and an avid student of languages. He learned Yiddish from his Russian parents. As an undergraduate at the University of Chicago, he studied German and French, and as a graduate and postgraduate at Yale University he concentrated on Nootka, a Canadian indigenous language. On his regular field trips through Canada, the United States, and Mexico, he collected data on more than 20 native languages. In Mexico, he developed programs for indigenous people to attain literacy in their own languages. Later, during World War II, he worked for the war department, editing dictionaries, providing linguistic analyses of foreign languages, and developing teaching materials for Spanish, Russian, Burmese, and Chinese.

Swadesh did his first theoretical work on phonemic analysis, i.e. the analysis of the sound structure or phonology of languages. Sapir, Leonard Bloomfield, Nikolai Trubetzkoy, and others had already advanced the concept of the 'phoneme', an abstract representation of sound types. Swadesh's contribution was to develop a set of principles to help the phonologist discover phonemes on the basis of the distribution of sounds in a given language. If, for example, one sound always occurs at the beginning of words, while another always occurs at the end, Swadesh suggested that these particular sounds are in 'complementary distribution' and thus could potentially be instances of the same sound type or phoneme. These principles were later applied to word and sentence elements by Zellig Harris. Distributional analysis thus became a general 'discovery procedure' for the basic elements of linguistic structure and has remained an integral part of linguistic methodology to this day.

In his extensive investigations of numerous languages, Swadesh gained an increasing appreciation of apparent lexical and structural similarities in different languages, and his interest in comparative historical linguistics grew. Since his study of nearly extinct languages was conducted with limited resources, he felt the need for a standardized procedure for quickly collecting crucial data yielding clues about language relationships.

To determine whether two given languages are related, historical linguists usually use the 'comparative method'. This means that they attempt to reconstruct an ancestor language on the basis of cognates, i.e. arguably related words from different languages (English *hound* is a cognate of German *Hund* 'dog'). Since cultural development is always accompanied by lexical innovations, Swadesh—and many linguists before him—felt that presumably more stable 'basic vocabulary' would be the best place to start looking for cognates. The basic vocabulary of a language describes body parts and functions, such as *skin, blood, drink*, natural phenomena like *water, sky, bird, smoke*, immediate sense experiences, and physical dimensions, such as *long, red, cold*. While the concept of basic vocabulary had been informally used before, Swadesh made it explicit by drafting a list of 100 word meanings that a fieldworker investigating any language could use for identifying the basic vocabulary of that language. The use of this list, now generally known as the 'Swadesh list', has drawn criticism from its inception. Many linguists believe that it is impossible to enumerate universal meanings, and that the identification of semantically equivalent words in different languages is often highly problematic. Nevertheless, the list has become a widely used tool in comparative linguistics.

Even though the notion of basic vocabulary was already contentious, Swadesh pressed forward and used his list for lexicostatistics, a quantitative method

1051

for measuring the similarity of languages. If the basic vocabulary of one language matches that of another by over 90%, Swadesh argued, these languages must be closely related. Most linguists believe that the reconstruction of ancestor languages provides more reliable evidence for the relatedness of languages than statistical analyses; hence, lexicostatistics continues to be viewed with suspicion. Yet, moderate linguists today concede that the Swadesh list and lexicostatistics may be useful for rough initial investigations or for situations where complete data are simply unavailable—which is, in fact, close to what Swadesh had in mind.

Even more controversially, Swadesh claimed that the 'decay' of basic vocabulary could be used for 'glottochronology', the dating of ancestor languages analogous to determining the age of fossils on the basis of radioactive decay. Swadesh came to believe that basic vocabulary decays at a rate of 14% over 1,000 years; hence, languages would retain on average about 86% of their basic vocabulary over this time span. Thus, if the basic vocabularies of the two related languages are found to match by 70%, they can be assumed to have developed from a single language that existed approximately 12 centuries ago. The assumption that basic vocabulary decay is generally uniform has been seriously challenged. If one allows that languages, just like societies, may develop at different rates at different times, the assumption of steady vocabulary decay in particular, and the glottochronological method in general, is seriously undermined.

Swadesh's name has remained symbolic for lexicostatistics and glottochronology, but his central place in the continuing and highly ideological debate about these and related issues seems to rest on misunderstandings and/or polemically slanted readings. He is, for example, accused of introducing lexicostatistics as a shortcut for investigation, attempting to avoid the hard work of reconstruction. However, Swadesh stated repeatedly that a detailed knowledge of the languages under investigation is crucial, and that other data must be considered. Likewise, Swadesh is misleadingly cited as a supporter of the theory that all languages have developed from a single ancestor (the 'monogenesis' theory). He certainly conceded that the instinctive vocalizations of early humanoids may be called a species-specific 'language', but he also surmised that as soon as arbitrary signs—i.e. true words and complex linguistic structures—entered the human repertoire, diversification was the instant result.

These ideas were the focus of his major book, *The origin and diversification of language*, but he died of a heart attack before he was able to complete it. His arguments thus remained somewhat sketchy, which may explain why Swadesh's even-handed and careful deliberations tend to be overshadowed by the bolder aspects of his thought.

Biography

Morris Swadesh was born in Holyoke, Massachusetts on January 22, 1909. He received his B.A. (1930) and M.A. (1931) for his dissertation on Nootka aspect, tutored by Edward Sapir, University of Chicago. He followed Sapir to Yale University, New Haven, Connecticut, for Ph.D. work on Nootka semantics, 1933, and then worked at Yale on synchronic phonological theory and on American English grammar in 1933–1937. He was Assistant Professor, University of Wisconsin, Madison in 1937–1939. He moved to Mexico City, where he became Director, Consejo de Lenguas Indígenas, and Director of Linguistics, Departamento de Asuntos Indígenas, in 1939; he was Professor, Instituto Politécnico Nacional de México, Escuela de Antropología, and Departamento de Asuntos Indigenas, in 1939–1941. Swadesh was a linguist for the War Department in New York City during World War II; he was Associate Professor, City University of New York in 1948, but lost his teaching appointment and had his passport revoked because of 'leftist' views and activities. He worked as a librarian at the Boas Collection, Library of the American Philosophical Society, Philadelphia, PA, 1949–1953, and carried out independent fieldwork in 1953–1956. He moved again to Mexico City, where he became Professor at the Instituto de Historia, Universidad Nacional Autónoma de México, and the Escuela Nacional de Antropología e Historia, 1956–1967. He was a member of the Linguistic Society of America (LSA) in 1931, Life Member in 1937; member of two special interest groups of the LSA in 1939; President of the Linguistic Section of the 29th International Congress of Americanists, 1939; and editor of *Word,* 1946–1949. Swadesh died in Mexico City on July 20, 1967.

References

Anttila and Raimo. 1972. An introduction to historical and comparative linguistics. New York: Macmillan; 2nd edition, as Historical and comparative linguistics, Amsterdam and Philadelphia: Benjamins, 1989.

Harris, Zellig. 1951. Methods in structural linguistics. Chicago: University of Chicago Press; as Structural Linguistics, 1960.

Hymes, Dell H. (ed.) 1964. Language in culture and society. New York: Harper & Row.

———. 1971. Morris Swadesh: from the first Yale School to world prehistory. The origin and diversification of language, by Morris Swadesh. Chicago: Aldine Atherton.

Lamb, Sidney M., and E. Douglas Mitchell. (eds.) 1991. Sprung from some common source: investigations into the

prehistory of languages. Stanford, CA: Stanford University Press.

Newman, Stanley. 1967. Morris Swadesh (1909–1967). Language 43. 948–57.

Sapir, Edward and Swadesh, Morris. 1932. The expression of the ending-point relation in English, French, and German, edited by Alice V. Morris, Linguistic Society of America, Language Monographs, Vol 10.

———. 1934. The phonemic principle. Language 10. 117–29.

———. 1937. A method for phonetic accuracy and speed. American Anthropologist 39(4). 728–32.

———. 1959. Linguistics as an instrument of prehistory. Southwestern Journal of Anthropology 15. 20–35.

———. 1971. The origin and diversification of language. Chicago: Aldine Atherton.

PHILIPP STRAZNY

See also **Haas, Mary Rosamond; Sapir, Edward**

Swedish and Scandinavian Languages

Scandinavian languages belong to the Indo-European North-Germanic languages. North-Germanic languages are spoken in the Nordic countries (Denmark (including The Faroes and Greenland), Finland, Iceland, Norway, and Sweden). The concept of Nordic languages (Swe. *nordiska språk*) covers all North-Germanic languages, and Scandinavian languages comprise the three languages Swedish, Danish, and Norwegian. The prehistoric period in Scandinavia is not well known, but by the second century CE, Ancient Scandinavian language(s) were spoken in Denmark, Southern and Central Sweden, and in southern Norway.

Based on older language history the North-Germanic languages have been divided into West-Nordic (Norwegian, Icelandic, and Faroese) and East-Nordic (Danish and Swedish). A more recent division separates North Scandinavian (Norwegian and Swedish) and South Scandinavian (Danish) from Insular Nordic (Icelandic and Faroese; cf. Torp 1982; Vikør 1993). The latter view is central here. The division into West and East Scandinavian was based on isoglosses, for example, the deletion of initial /j/ in West Scandinavian (=Old Norse): Nynorsk *ek*, and Faroese *eg*, Swedish *jag* and Danish *jeg*, 'I'. Changes during the Middle ages in Norwegian/Swedish vs. Danish are considered more significant than the West – East split (in the sixth – seventh centuries CE). For example, the four-case system in Icelandic (with Nominative, Genitive, Dative, and Accusative) developed into a two-case system in the Scandinavian languages (Nom. and Gen.). Another change was the loss of the synthetic verb conjugation, which in the Old Norse system expressed person, number, and mood. Suffixed definite articles developed from the demonstratives, and the passive voice, which evolved from the reflexive pronoun *sik* 'self', for example, Swe. *kalla* 'call' vs. *kalla-s* 'be called', are characteristic of mainland Scandinavian.

Within phonology, the syncope (deletion between consonants) of short, unstressed vowels in plurisyllabic words simplified the Ancient Scandinavian syllable structure. For example, the inscription in the fifth century golden horn from Danish Gallehus, *ek hlewagastiR holtijaR horna tawido*, 'I, HlewagastiR from Holt, made the horn', results in the reconstructed forms: **ek hlégestr hyltir horn táða*, in Old Norse. Plurisyllabic words like *Hlégestr* 'dropped' syllables. Deletion of Ancient Scandinavian final –n caused a changed infinitive form: *fullian* > Swe. *fylla* 'fill', *lausian* > Da. *løysa* 'loosen'. The Umlaut (common Germanic vowel shift) complicated the morphology structure from Ancient Scandinavian to Old Norse. Particularly, /i/ in suffices influenced /a/, /o/, and /u/ in word roots: *ansiR* > *æsir* 'pagan gods', *fōtiR* > *fØtr*, 'feet'. In words like *mann ~ menn* 'man ~ men' and *mús ~ mýss* 'mouse ~ mice', the vowel shift received phonemic weight: Swe. singular *man* vs. plural *män*. The three Scandinavian languages (and German) share the development of /æ/, /Ø/, and /y/, which are represented in writing (Swe. *ä* and No./Da. *æ* for /æ/, Swe. *ö* and No./Da. *ø* for /Ø/ and y for /y/). Swedish also has *å*, for /o/, which was recently adopted by Norwegian and Danish for the digraph *aa*. Diphthongs changed to monophthongs: *lausian* > Swe. *lösa* 'loosen'. The change of diphthongs to monophthongs did not affect Old Norse or some Norwegian dialects: Nynorsk has *stein*, 'stone', *laus* 'loose', and *Øy* 'island', compared to Swedish/Danish *sten*, *lös/løs*, and *ö/ø*.

Danish vs. North Scandinavian Languages

Three changes initiated a unique development for Danish. The neutralization of unstressed vowels in inflectional endings in Danish developed a central vowel/schwaa, which is still spelled with an –e: Da. *holde*, Swe. *hålla* 'to hold'. This initiated a reduction of inflectional morphology and its functions.

Danish unvoiced stops after long vowels became weaker pronounced (lenition). The older forms *rep*

'rope', *bita* 'bite', *høk* 'hawk', and *leka* 'play' developed into *reb*, *bide*, *høg*, and *lege.* This has not been consequently reflected in later spelling conventions in Danish. In Danish, the voiced stops shifted to fricatives. The adjacent dialect areas in Southern Sweden and Norway partly followed suit, but they did not introduce the full set of fricatives.

Danish developed a glottal stop, which was paralleled by the development of a tonal system in Norwegian and Swedish, with two tonemes. The tones make a difference between words such as Swe. ´*anden* 'the duck' (Swe. akut accent) and ˣ*anden* 'the spirit' (grav accent). The glottal stop is not reflected in Danish spelling; it is understood as a prosodic, not a segmental feature. Insular Nordic has neither the glottal stop nor the toneme system. These three phonetical changes added to the differences between Danish and Swedish/Norwegian; Danes and Swedes today have difficulties understanding each other.

Swedish

The separate Swedish development can be divided into five periods: Runic Swedish 800–1225 CE, Classical Old Swedish 1225–1375, Late Old Swedish 1375–1526, Early Modern Swedish 1526–1732, and Modern Swedish from 1732–. These periods are based on features of the written language. The Runic inscriptions (Germanic by origin) were possibly used as early as the second century CE in Denmark. Runes were used in carvings in hard material such as wood, bone, and stone; they have been well retained in runic stones. Runes were used for almost a millennium, even for everyday purposes in the late Middle Ages.

The Scandinavian languages were stabilized by the translations of the New Testament (in 1526 to Swedish) and the Bible (in 1541, Gustav Vasa's Bible). Printing had been introduced in 1483. Some of the major typological changes of Swedish from a synthetic to a more analytic language were completed by the end of the fifteenth century. In 1732, the newpaper *Then Swenska Argus* started appearing (similar to the English *The Spectator*), which had a modernizing impact on the writing of Swedish.

The administration of the Swedish kingdom was centralized early and efficiently by the Vasa dynasty (1523–1654). It used a Stockholm-based chancery style, which facilitated the standardization of Swedish and the evolution of a spoken standard Swedish. The cultivation of Swedish was supported by the foundation of the Swedish Academy in 1786 by King Gustav III. During this period, French also had a great influence on the court and its language.

The Swedish dialects are divided into six major groups: South Swedish (Scania and adjoining provinces), Guthnic dialects (Northwest of Scania;

Western Sweden), Central Swedish, North Swedish, Gutnish (Gotland), and East Swedish dialects (in Finland, influenced by Finnish phonology). The differences between the peripheral dialects are considerable.

Personal pronouns are compulsory to indicate subject person (only tense is expressed in the verbs): *jag talar* 'I speak', *du talar* 'you speak' (second-person singular), *han talar* 'he speaks', *vi talar* 'we speak'. Regular (*titta–tittade–tittat* 'look, looked, looked') and irregular forms (*springa–sprang–sprungit* 'run, ran, run') of verbs exist. The definite–indefinite article system is complex, since congruence of definiteness, gender, and number is expressed. Swedish (and Danish) only has Neuter (de**t**-genus) and Uter (de**n**-genus) gender. For example, *det stora huset* 'the big house', *ett stort hus* 'a big house', and, *den gamla staden* 'the old town', *en gammal stad* 'an old town'.

Word order (basically SVO) has syntactic functions, for example, indicating the type of clause (main–subordinate; declarative–question): *Han <u>kommer inte</u>* (main clause) 'He will not come', vs. *Han säger, att han <u>inte kommer</u>* 'He says, that he will not come'; *<u>Han kommer</u> snart* 'he will come soon', vs. *<u>Kommer han</u> snart?* 'Does he come soon?'.

Swedish is the de facto official language of Sweden. About 8.2–8.3 million speakers of the nine million inhabitants speak it as a first language; the remainder speak it as a second language. In Finland, Swedish is an official national language, with about 300,000 speakers. Emigration to North America started already in the seventeenth century, but had its peak in the late 1800s. The Swedish targets were mainly the farming areas around the Great Lakes. About one million people left Sweden for North America; about 70,000 speak Swedish today (the United States of America). Modern trans-Atlantic migration has targeted the West and East coasts.

Danish

Danish and Swedish have experienced similar developments for their written languages. Old Danish is set to 800–1100 CE, Early Middle Danish 1100–1350, Late Middle Danish 1350–1525, Early Modern Danish 1525–1700, and Modern Danish 1700–. Christianity had a clear impact on this process, both during the eleventh century, and with the Lutheran Reformation in the sixteenth century. Low German language contact effects were extensive during the Late Middle Danish period.

Danish Vikings created a strong position for Danish in Northern Europe. It has retained a dominant position vis-á-vis Norwegian, Icelandic, and Faroese (and Greenlandic). Danish was an urban high-status variety in Norway for seven to eight centuries. This has brought the Norwegian vocabulary close to Danish.

Danish has itself been under continuous pressure from German and Latin.

Similar to Swedish, the oldest written sources, in addition to runic inscriptions, are regional law texts from the thirteenth century. Beginning in the fourteenth century, legends, chronicles, and religious literature were written in and translated into Danish. The influential Bible of King Christian III in 1550 was to a large extent based on Danish dialects. From that period onward, the Zealand dialects, including Roskilde and Copenhagen, have been important. The standardization process was largely adapted to Copenhagen upper-class people's style. The German Oldenburg monarchy, which ruled the country for four centuries up to the mid-nineteenth century, preferred German. Also, Latin and later French were prestigious languages, whereas Danish remained the 'people's' language until the eighteenth century. With the writer Ludvig Holberg (1684–1754), a stable literary development started, which pointed out the route to present-day standard Danish.

Three main Danish dialect areas are recognized: Jutish in Jutland, Insular Danish in Funen and Zealand, and 'East-Danish', in Bornholm and the lost provinces in Southern Sweden. Similar to Swedish, an extensive dialect leveling has taken place recently.

Danish is the first language of about 4.8 million of the 5.1 million inhabitants of Denmark. It is the official language of Denmark, and has a secondary official status in the Faroes and Greenland. About 45,000 speak Danish in North Schleswig, where it can be studied at all levels in school. Danish emigration to North America was extensive in the nineteenth and twentieth centuries. Today, 35,000 descendants speak the language (United States of America), many of whom live on the West Coast.

Norwegian

Norwegian stems from Old Norse (West Scandinavian), 700–1350 CE. The Middle Norwegian period is set to 1350–1525, and Modern Norwegian 1525–. Old Norse represents a more synthetic and Modern Norwegian represents an analytic stage of the language. Written forms of Old Norse include the Icelandic Sagas. The historian Snorri Sturlasson (1179–1241) termed the Old Norse of his day Danish. The idea of different languages (Danish, Icelandic, Norwegian) has thus been promoted later.

Middle Norwegian was under heavy pressure from Danish. Written Norwegian was practically replaced by Danish writing by 1525. Further, the Danish Bible of Christian III was used in Norway. This made the attempt to create a stable, written Norwegian unsuccessful. A Danish-based high variety developed in the towns.

After the Danish secession (1814) of Norway, Danish dominance was challenged. It remained a highly central issue during the whole period of the Norwegian–Swedish Union, 1814–1905. One language political direction was headed by the self-taught linguist Ivar Aasen (1813–1896), who undertook a study of south and west-Norwegian dialects. He founded modern Norwegian dialectology, and through his grammar (1864) and dictionary (1873) based on these dialects, he created a standard written Norwegian, *Nynorsk*. This was in line with the ideas of the nationalistic Romantic era. Nynorsk was given official recognition in Parliament in 1885. The support for Nynorsk has decreased steadily since the 1940s.

Another approach was more pragmatic. The teacher Knud Knudsen (1812–1895) Norwegianized the established, written, Danish-based language according to the middle-class townspeople's speech habits. These were influenced by southeastern dialects and Danish. This created the second Norwegian standard, *Bokmål* (or Dano-Norwegian). Through two spelling reforms (1907 and 1917), Bokmål became more adapted to Norwegian. After the first translation from Danish to Bokmål in 1919, Bokmål has been considered a separate language from Danish. About 1880, the two forms of Norwegian were connected to the political struggle between on the one hand a more radical, parliamentary wing—supporting Nynorsk—and a more urban, Oslo-based conservative wing—supporting Bokmål.

Since the linguistic differences were small it was suggested already in the 1880s, that the two standards be amalgamated to a common Norwegian (*Samnorsk*). The idea slowly gained ground, but after fierce opposition during the 1960s, the official support for this view has been downgraded. A natural leveling has partially closed the linguistic gap, but attitudinal resistance is still strong on both sides. Bokmål is used by more than eight out of ten Norwegians as their school language and written code.

Compared to Danish and Swedish language policies, Norwegian official policy has been both more liberal in relation to dialects, and more puristic regarding language contact effects (from English). Norwegian purism is less protective than the Icelandic attempts to fully replace the influx of international loanwords.

Norwegian dialects are divided into five groups: Western, Central (covering the inland valleys), Eastern, Trønder, and Northern dialects. Nynorsk and its base dialects (Western and Central), have greater similarities with the Insular Nordic languages.

About 95% of the Norwegians speak the official Nynorsk or Bokmål as their mother tongues. About 800,000 Norwegians emigrated to North America. Today, 80,000 know Norwegian in the United States. Many of them migrated to the shores of the Great Lakes, as did the Swedes.

Linguistic Examples of Swedish, Danish, and Norwegian

	Swedish	Danish	Norwegian	
			Bokmål	*Nynorsk*
'one'	ett	et	et	eit
'eye'	öga	øje	øye	auga
'weak'	svag	blød	bløt	blaut
'week'	vecka	uge	uke	veke
'I'	jag	jeg	jeg	eg
'she'	hon	hun	hun	ho
'we'	vi	vi	vi	vi, me
'who'	vem	hvem	hvem	kven
'behind'	bakom	bagved	bakom	attom
'not'	inte	ikke	ikke	ikkje
'fifty'	femtio	halvtres	femti	femti
'swear'	svära	bande	banne	banne
'count'	räkna	tælle	telle	telje
'window'	fönster	vindue	vindu	vindu, vindauge

References

Barðdal, Jóhanna, Nils Jörgensen, Gorm Larsen, and Bente Martinussen. 1997. Nordiska - Våra språk förr och nu (Nordic—our languages in the past and present). Lund: Studentlitteratur.

Haugen, Einar. 1976. The Scandinavian languages: an introduction to their history. London: Faber & Faber Ltd.

Klintborg, Staffan. 1999. The transience of American Swedish, Lund Studies in English, 98. Lund: Lund University Press.

Molde, Bertil, and Allan Karker (eds.) 1983. Språkene i Norden (The languages in the Nordic countries). Stockholm: Esselte, Copenhagen: Gyldendal and Oslo: Cappelen.

Torp, Arne. 1982. Norsk og nordisk før og nå (Norwegian and Nordic languages in the past and present). Oslo: Universitetsforlaget.

Vikør, Lars. 1993. The Nordic languages: their status and interrelations. Oslo: Novus Press.

JARMO LAINIO

See also **Celtic Languages; Finnish and Finnic Languages; German; Old English; Scandinavia**

Sweet, Henry

There were three significant phonetician-linguists in Britain in the nineteenth century: Alexander Melville Bell (1819–1905), Alexander John Ellis (1814–1890), and Henry Sweet (1845–1912). Bell is remembered primarily for his 'Visible Speech' alphabet and the theory of general phonetics that underlies it; Ellis for his studies of English pronunciation, both synchronic and diachronic; and Sweet for his erudite presentation, in a consistently structured and intelligible manner, of material on phonetics and various forms of medieval English. Of the three, it is Sweet who has exerted the strongest influence over the direction that the subjects have taken over the past 130 or so years.

Sweet's interests in the older Germanic languages began relatively early. One of his schoolteachers (and an Anglo-Saxonist), Thomas Oswald Cockayne (1807–1873), may have been instrumental in introducing him to the two languages. On leaving school, he undertook a year's study of Comparative and Germanic Philology at Heidelberg University. A consequence was that he began to collect materials during his teenage years for what was later to become his *Student's dictionary of Anglo-Saxon* (1897). While he was still an undergraduate at Oxford, his edition of King Alfred's *Pastoral Care* was published.

He read Classics ('Greats') at Oxford, but he was hardly assiduous, spending most of his time working, virtually alone, in the field of Germanic languages. At the time of his graduation, he had published not only the *Pastoral Care*, but had critically reviewed a number of scholarly works in the academic press, and had read three papers to the Philological Society. Within a year of graduating, he published his *History of English sounds* (1874).

Throughout his life, he maintained a wide spectrum of academic interests, including phonetics, spelling reform, shorthand, grammar, the teaching and learning of languages, general linguistics, the history of English and other Germanic languages, and literature. Since he had to fulfill two roles, that of the scholar and, secondly, the private teacher whose books for undergraduates and other learners would generate a certain amount of income, he adopted the general strategy of publishing advanced material first, then simplifying some of it in works written intentionally for the student learner.

Meetings of the Philological Society in London were the locus for the exposition of many of his views on language and languages, and he read papers there on a variety of subjects, especially the contemporary forms of a number of languages, including Danish, German, Icelandic, Irish Gaelic, and Swedish. In his

later years, his field of interest widened to take account of aspects of the linguistic structure of Arabic, Chinese, Norwegian, Portuguese, Sanskrit, and Welsh.

His *Handbook of phonetics* (1877) inspired a whole tradition of phonetic studies, especially in Britain and Europe. It expounds a general phonetic theory, illustrated by phonetic transcriptions of various languages, and concludes with an excursus on the nature of phonetic notational systems, including a pivotal discussion of Broad and Narrow Romic notations—equivalent in most respects to phonemic and allophonic notations. Two works on phonetics for the student learner, the *Primer of spoken English* (1890) and the *Primer of phonetics* (1890), present a less detailed account of phonetic theory. The text of the *Primer of spoken English* is essentially a translation of another of Sweet's influential works, his *Elementarbuch des gesprochenen Englisch* (1885), written for German learners of English at a time when radical changes were under way in European schools and universities in the teaching and learning of modern languages. A further work on English phonetics, the *Sounds of English*, appeared in 1908. *The practical study of languages* (1899) discusses in detail the theories and practices of teaching and learning languages.

Sweet's ideas about grammar and grammatical analysis developed relatively slowly. His *New English grammar* (1892, 1898) deals with the contemporary and historical forms of the language. Two simplified accounts, based on the work, are his *Short historical English grammar* (1892) and the *Primer of historical grammar* (1902).

The edition of the *Epinal glossary* (1883), together with a companion volume, *The oldest English texts* (1886) and *King Alfred's Orosius* (1883), provided scholars with the necessary tools for advancing the critical study of the earliest periods of Old English.

The series of works on earlier forms of English, written expressly for the novice learner rather than the scholar, began with the *Anglo-Saxon reader* of 1876, followed by a less demanding account, the *Anglo-Saxon primer* in 1882. Three further intermediate textbooks for students of Old English were Ælfric's *Homilies* (1885), *Extracts from Alfred's Orosius* (1885), and the *Second Anglo-Saxon reader* (1887). For students embarking on the study of Middle English, there were two *Middle English primers* (1884, 1886).

His Icelandic *Primer* (1886) is a classic example of his ability to present the essential facts of a medieval language in a form that is immediately attractive to the learner. The essential linguistic information is described synchronically, with no diachronic or cognate Germanic materials to distract the learner. Sweet's *History of language* (1900), written for the general reader, is an exposition of the principles that underlie comparative philology, with particular reference to the Neo-Grammarian thesis on language change.

Biography

Henry Sweet was born in London on September 15, 1845. He was educated at private schools, but primarily at King's College School, London. He received his B.A. in 1873 from Oxford University, Ph.D. (Hon Causa) from Heidelberg University in 1875; and LL.D. (Hon Causa) from Glasgow University in 1890. He studied Comparative and Germanic Philology at Heidelberg University, 1863–1864. He was an undergraduate student at Balliol College, Oxford, 1869–1873. He obtained his only full-time teaching post, Reader in Phonetics, at Oxford University in 1901. He was President of the International Phonetic Association from 1887 until his death. He was Corresponding Member of the Munich Academy of Sciences, and of the Royal Prussian Academy of Sciences, and an Ordinary Member of the Royal Danish Academy. Sweet died in Oxford on April 30, 1912. The character of Henry Higgins in George Bernard Shaw's (1856–1950) *Pygmalion* is based to some extent on that of Sweet.

References

Henderson, E.J.A. (ed.) 1971. The indispensable foundation. A selection from the writings of Henry Sweet. London: Oxford University Press.

MacMahon, M.K.C. 2001. Modern language instruction and phonetics in the later nineteenth century. Geschichte der Sprachwissenschaften, ed. by S. Auroux et al. Berlin and New York: Walter de Gruyter.

Sweet, H. 1871. King Alfred's West-Saxon version of Gregory's pastoral care. With an English translation, the Latin text, notes, and an introduction. London: Trübner & Co. [published 1872].

———. 1874. A history of English sounds from the Earliest Period, including an investigation of the general laws of sound change, and full word-lists. London: Trübner & Co.

———. 1876. An Anglo-Saxon reader in prose and verse, with grammatical introduction, notes, and glossary. Oxford: Clarendon Press.

———. 1877. A handbook of phonetics, including a popular exposition of the principles of spelling reform. Oxford: Clarendon Press.

———. 1882. An Anglo-Saxon primer with grammar, notes and glossary. Oxford: Clarendon Press.

———. 1883. The Epinal glossary, Latin and Old-English of the eighth century. Photo-lithographed from the original MS … with transliteration, introduction, and notes. London: Trübner & Co.

———. 1885. Elementarbuch des gesprochenen Englisch Grammatik, Texte und Glossar. Oxford: Clarendon Press.

———. 1886. An Icelandic primer, with grammar, notes and glossary. Oxford: Clarendon Press.

———. 1886. The oldest English texts, edited with introductions and a glossary. London: Trübner & Co.

———. 1888. A history of English sounds from the Earliest Period, with full word-lists. Oxford: Clarendon Press.

SWEET, HENRY

———. 1890. A primer of phonetics. Oxford: Clarendon Press.
———. 1890. A primer of spoken English, Oxford: Clarendon Press.
———. 1892/1898. A new English grammar, logical and historical. Part I: Introduction, phonology, and accidence (1892); Part II: Syntax (1898). Oxford: Clarendon Press.
———. 1897. The student's dictionary of Anglo-Saxon. Oxford: Clarendon Press.

———. 1899. The practical study of languages. A guide for teachers and learners. London: J.M. Dent & Co.
———. 1900. The history of language. An introduction to the principles of comparative philology. London: J.M. Dent & Co.
———. 1908. The sounds of English. An introduction to phonetics. Oxford: Clarendon Press.
Wrenn, C.L. 1946. Henry Sweet. Trans Philological Society.

MICHAEL K.C. MACMAHON

Switch-Reference

Switch-reference (first described by William Jacobsen in 'Switch-reference in Hokan-Coahuiltecan', 1967) is one of several subject-linked reference tracking mechanisms that apply outside the domain of a single clause. Switch-reference markers indicate whether the referent of an argument (typically the subject) of one clause is the same as or different from the referent of an argument (typically the subject) of another clause. Switch-reference systems are found in a wide range of languages in North America, Australia, Papua New Guinea, North Asia, and Africa.

Typical Switch-Reference Systems

In a typical switch-reference system, the verb of one clause, the marking clause, is marked to indicate whether or not its subject has the same referent as the subject of another clause, the reference clause. The marking clause can be subordinate to the reference clause, as in the Maricopa examples (1ab) or coordinate with the reference clause as in the Chickasaw examples (2ab). Languages limit what syntactic relationships hold between marking and reference clauses: in some languages, switch-reference is used across a wide range of coordination and subordination relationships, while in others it is used in more limited structures. Switch-reference applies only between clauses that combine to form a single syntactic sentence.

(1a) paan nya-m-chew-<u>k</u>   m-maa-uum
       bread when-you-make-ss you-eat-INCOMPLETE
       'When you make the bread, you'll eat it'
(1b) paan nya-m-chew-<u>m</u>   '-maa-uum
       bread when-you-make-ds I-eat-INCOMPLETE
       'When you make the bread, I'll eat it'
(2a) hi'lha-<u>cha</u> talowa
       dance-COORDINATE = ss sing
       'He danced and sang'
(2b) hi'lha-<u>na</u> talowa
       dance-COORDINATE = DS sing
       'He danced and he (someone else) sang'

(Chickasaw data are from Munro, P. 'When 'Same' is not 'Not Different', 1983: 223)

In (1a) and (2a), the underlined morphemes labeled 'ss' show that the subject of the marking clause has the same referent as the subject of the reference clause; in (1b) and (2b), the morphemes labeled 'DS' show that the subjects of the marking clause and the reference clause have different referents.

Switch-reference marking obligatorily tracks the reference of subjects across clauses within a sentence, even when other morphology makes the switch-reference marking redundant. For example, the agreement prefixes on the verbs in (1b) show that the marking clause has a second-person subject while the reference clause has a first-person subject; but a different-subject suffix is still required on the verb of the marking clause. Similarly, the agreement prefixes on the verbs in (1a) show that both clauses have first-person subjects, but the same-subject suffix is still required.

More complex examples show that the switch-reference marking is sensitive to the syntactic subject, rather than to a logical or underlying subject or agent.

When one clause is subordinate to the other, the marking clause is always the subordinate one, regardless of linear order. For example, in Maricopa, a subordinate marking clause can be initial, final, or embedded in the middle of its reference clause.

Typical Variation in Switch-Reference Systems

Individual languages deal differently with a number of predictable structures, including (1) the assignment of switch-reference marking when the referents of the subjects of the relevant clauses are overlapping, rather than entirely different or completely identical; (2) the assignment of switch-reference marking when the subject of one or both of the clauses does not have a referent; (3) the combination of switch-reference markers with other functional markers (for example,

in Chickasaw (2a), *-cha* marks both coordinate and same subject); and (4) the use of switch-reference-marked verbs that appear to have a primarily discourse function as transitions between sentences or paragraphs.

Uncommon Variants in Switch-Reference Systems

In some languages (e.g. Pima), switch-reference markers are not verbal affixes, but independent particles. In other languages, the system is extended from subjects to include other arguments. For example, Peter Austin (in 'Switch-reference in Australia' in *Language* 57 (1981):325) describes Warlpiri as having a three-way contrast: if the subordinate clause subject is coreferential with the subject of the main clause, the subordinate clause verb is marked with *-karra*; if it is coreferential with the object or dative of the main clause, the marker is *-kurra*; otherwise the marker is *-ngkarni/-rlarni.* In still other languages, the switch-reference system is restricted to a subset of persons, or interacts with a person hierarchy as in Quechua.

Other Reference Tracking Systems

Switch-reference must be distinguished from a number of other systems used by languages to indicate identity or lack of identity of reference of arguments across clauses. For example, a number of African languages have a 'logophoric' pronoun or verbal affix, which appears in clauses that report someone's ideas, words, or feelings. The use of the logophoric pronoun in the reported clause indicates that its referent is the same as the individual whose ideas, words, or feelings are being reported. The logophoric form and the argument it refers back to are not necessarily limited to the subject position and it is limited to reported contexts. Switch-reference, in contrast, seems to be much more strictly linked to syntactic structure and much less linked to particular semantic or discourse contexts; unlike logophoric forms, switch-reference markers are never pronouns and typically are not part of the ordinary verb-agreement system.

Analyses

There are two basic approaches to switch-reference: syntactic/semantic and discourse/pragmatic.

Syntactic/semantic approaches, like Daniel Finer's 1985 *The formal grammar of switch-reference* and George Broadwell's 1997 article in *Atomism and binding,* analyze the assignment or interpretation of switch-reference markers as being controlled by the syntactic structure of the sentence. These accounts take the assignment or interpretation of reference of subjects as basic to the structure, with other functions derivative from this primary use.

Discourse/pragmatic approaches, like David Rising's 1992 *Switch reference in Koasati discourse* and Yan Huang's 2000 *Anaphora: A cross-linguistic study*, assume that the primary use of switch-reference markers is to mark topic or other discourse continuity and the marking of subject identity or lack of identity is derivative.

References

Austin, Peter. 1981. Switch-reference in Australia. Language 57.

Broadwell, George A. 1997. Binding theory and switch-reference. Atomism and binding, ed. by Hans Bennis, Pierre Pica, and Johan Rooryck. Dordrecht: Foris Publications.

Finer, Daniel. 1985. The formal grammar of switch-reference. New York and London: Garland Publishing, Inc.

Haiman, John, and Pamela Munro (eds.) 1983. Switch-reference and universal grammar. Amsterdam: John Benjamins.

Huang, Yan. 2000. Anaphora: a cross-linguistic study, Oxford Studies in Typology and Linguistic Theory. Oxford: Oxford University Press.

Jacobsen Jr., William. 1967. Switch-reference in Hokan-Coahuiltecan. Studies in Southwestern ethnolinguistics: meaning and history in the languages of the American Southwest, ed. by Dell H. Hymes and William E. Bittle. The Hague and Paris: Mouton & Co.

Munro, Pamela. 1983. When 'same' is not 'not different'. In Haiman and Munro.

Rising, David P. 1992. Switch reference in Koasati discourse. Dallas: Summer Institute of Linguistics and the University of Texas at Arlington.

Stirling, Leslie. 1993. Switch-reference and discourse representation theory. Cambridge: Cambridge University Press.

LYNN GORDON

Switzerland

The Swiss Confederation consists of 26 cantons (states, provinces) within an area of only 41,284 square kilometers. According to the 1990 census, this small country has a population of 6,873,867 speaking four languages: German, French, and Italian, which are shared with its neighbouring countries, and

Romansh, a distinct Romance language, that is, one descended from Latin and related to French, Italian, Spanish, and others. This linguistic mix is the result of the historical processes leading to the composition of modern Switzerland. The original confederation was an oath of eternal allegiance taken between four German-speaking cantons in the mountainous centre of the country, occurring in 1291. The German language had been brought to this area by the Alemannic invaders, part of the expansion of Germanic peoples into the old domains of the Roman Empire from the fifth century onward. After 1291, the confederation expanded, adding more cantons, reaching its present composition only at the beginning of the nineteenth century. While some of these cantons were, like the original four, German speaking, those in the west spoke French, while the incorporation of the southern areas brought in Italian and Romansh. The linguistic boundaries shown on the map below are in fact older than the state boundary of Switzerland. Even the increasing Germanisation of the Romansh area is a process that began well before the area's incorporation into the Swiss Confederation.

All four of these languages are defined as 'national languages' in the federal constitution, and since 1996, when Romansh was added to the other three—as 'official languages' of the confederation. Multiple national languages are unusual, especially for Europe, and Switzerland has been cited as a possible model for the European Union with respect to the accommodation of several official languages. Although Switzerland is at a national level quadrilingual, it is not the case that the four languages are evenly distributed through the population, or that all areas of Switzerland are multilingual. In practice, despite much internal migration and the large foreign population in Switzerland, most areas of the country are essentially monolingual. Each of the three major languages has its territory in which it functions as the single language of public life. The Swiss school system aims to provide all Swiss with some knowledge of a second national language; however, this goal is by no means fully achieved. The following statistics and map show the numerical and geographic distribution of the country's four languages.

Numbers and percentages of Swiss population who nominated one of the national languages as their main language in the 1990 census:

German:	4,374,694	63.6%
French:	1,321,695	19.2%
Italian:	524,116	7.6%
Romansh:	39,632	0.57%

Territories of the four languages of Switzerland (1: French; 2: German; 3. Italian; and 4: Romansh):
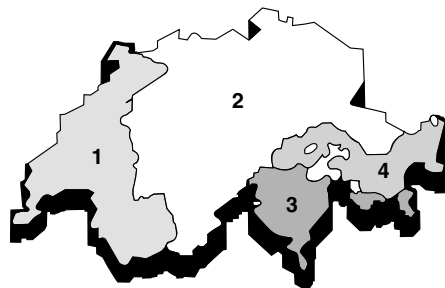
The patterns of language use in communication among the language communities of Switzerland vary widely, with the Italian and Romansh speakers far more likely to adapt to their compatriots from the larger communities. Indeed, English—a popular choice as the third school language—is now widely used as a lingua franca. The situation is exacerbated by the perceived dominant position of the German-speaking community, and by the existence of diglossia in the German-speaking area.

Diglossia, as originally defined, is the existence of two distinct forms of the same language within the same community, each with its own functions. Ferguson's (1959) original article on this phenomenon uses German-speaking Switzerland as one of the canonical examples. In this area, Swiss Standard German, which is very similar but not identical to Standard German, is used for most written functions and such highly formal spoken genres as university lectures, while local dialects, known collectively as *Schwyzertüütsch,* are used for most spoken purposes. These dialects are so different from Standard German as to be practically mutually unintelligible with it. The differences are primarily due to varied histories of their sound and grammar systems, but there are also lexical items unique to the Swiss dialects. The table below presents a few examples of these characteristic phonological and lexical distinctions:

English	Zurich German	Standard German
newspaper	*Zytig*	*Zeitung*
house	*Huus*	*Haus*
to take out	*uusenee*	*herausnehmen*
let's go home	*Gömer häi*	*Gehen wir nach Hause*
listen	*lose*	*zuhören*
one, two, three, four, five	*äis, zwoo, drüü, vier, föif*	*eins, zwei, drei, vier, fünf*
I was there	*I bi dëet gsy*	*ich war dort*

There is widespread reluctance among German-speaking Swiss, and in many cases inability, to use Standard German easily for spoken purposes, especially for informal ones; yet, Standard German is what other Swiss learn at school.

Historically, the French and Italian areas of Switzerland had local dialects, but the situation in both these areas differs from that found in German-speaking Switzerland. The local dialects of French belong to the Franco-Provençal group, a family of dialects sharing language attributes with both northern French (to which Standard French belongs historically) and the Occitan languages of the south of France. However, to all intents and purposes, these dialects are now lost in Switzerland. With the exception of a few very old people, everyone in the French-speaking parts of Switzerland speaks Standard French. Standard French is not identical to Parisian French; rather, it incorporates a few older features that have been replaced in Parisian French, as demonstrated in the table below:

English	Swiss Standard French	Parisian French
he must be helped	*il faut **lui** aider*	*il faut **l'**aider*
seventy	*septante*	*soixante-dix* (sixty plus ten)
eighty	*huitante*	*quatre-vingts* (four twenties)
ninety	*nonante*	*quatre-vingt-dix* (four twenties plus ten)

The local dialects of Italian are still spoken, but until recently had low status. Standard Italian was regarded as the language of prestige. In recent decades, however, the Swiss dialects of Italian have gained in usage and prominence as a mark of local identity.

Romansh has been spoken continuously in the southeast portions of Switzerland since Roman times. Linguistically, it shares features with both northern Italian dialects and French. Like other languages of Switzerland, particularly German, it exists as a chain of dialects that have been standardised in various forms since the time of the Reformation in the sixteenth century. Of these five written forms, those of the upper Rhine valley and the lower Engadine (the left- and right-hand ends of area number 4 on the map) have enjoyed particular support. However, despite its long history and its official status at both federal and cantonal levels, Romansh is endangered. A steadily decreasing proportion of Swiss report Romansh as their first language in the

census, held every ten years, and Romansh's territory has been shrinking under pressure from German for many centuries. Virtually all speakers of Romansh are fluent in German, often in addition to other languages as well. There have been and continue to be, however, strenuous efforts to ensure the survival of Romansh.

The precarious state of the language is not helped by the existence of the five written versions. In 1982, a unified written form of the language was introduced greatly facilitating the publication of official materials in Romansh. Even works of literature are now being written in this standard. Founded in 1919 as an umbrella organization for the growing number of societies and groups founded in the previous century to promote and protect Romansh, the Lia Rumantscha, with its seat in Chur, is the main organisation coordinating and driving the work of language maintenance and language planning for Romansh. Although not a statutory body, the Lia Rumantscha is financed by very generous grants from both federal and cantonal governments. Its activities include translation services, terminology creation and definition, publication, and language promotion. In those municipalities with a majority of Romansh speakers, Romansh is used as the language of instruction at junior levels of school, and is available as a subject at secondary and even tertiary levels of education.

The linguistic situation in Switzerland is complicated, and is a matter of intense discussion within the country itself, not least because of the perception that the linguistic communities of Switzerland do not so much live together as past one another. Pride in this 'diversity in unity', however, is an important feature of Swiss identity.

References

Clyne, Michael. 1995. The German language in a changing Europe. Cambridge and New York: Cambridge University Press.

Dürmüller, Urs. 1997. Changing patterns of multilingualism: from quadrilingual to multilingual Switzerland. Zurich: Pro Helvetia.

Duval-Valentin, Marianne. 1983. La situation linguistique en Suisse. Language reform: history and future, Vol. 1, ed. by István Fodor and Claude Hagège, Hamburg: Buske.

Ferguson, C.A. 1959. Diglossia. Word 15. 325–44.

Watts, Richard J. 1991. Linguistic minorities and language conflict in Europe: learning from the Swiss experience. A language policy for the European community, ed. by Florian Coulmas. Berlin and New York: Mouton.

The activities of the Lia rumantscha are described at <http://www.liarumantscha.ch>.

RAY HARLOW

Syllable: Structure

The syllable is defined as a unit of cognitive organization, comprised of a 'sonority peak'. Sonority is traditionally defined as *intrinsic loudness* (Chomsky and Halle 1968). It is rather unclear as to what sonority might translate into in verifiable phonetic terms, but it is generally accepted that the major phonological classes can be ranked from high to low sonority as follows: vowels, glides, liquids, nasals, and obstruents. Following researchers such as Itô, (1989), the segments of a word are exhaustively parsed into syllables. This is termed 'syllabic licensing'. For example, marking syllable boundaries with parentheses, a word like *system* is syllabified as follows: [(sɪs)(təm)].

It is a question of some interest whether syllables are further divided into smaller constituents above the segment level. Many researchers assume that a syllable is subdivided into *onset* and *rhyme* constituents, where the rhyme constituent is perhaps further subdivided into a *nucleus* and *coda*. The onset is composed of all the segments that precede the peak of the syllable. The nucleus is composed of all the vocalic material of the peak; all the material to the right of the nucleus is in the coda. An English syllable like *trounce* [trawns] would have the following structure:
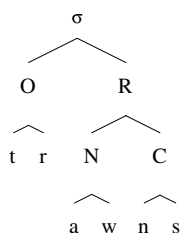
The subsyllabic constituents presented are abbreviated in the obvious fashion and a Greek sigma 'σ' is used for a syllable.

Evidence for the onset-rhyme division comes from two main sources: phonotactics and stress. First, in many languages, while there are a number of restrictions on what consonants can occur together in an onset, there are very few restrictions on what segments can co-occur together across the onset-rhyme division. In English, for example, the only known restriction that plausibly spans the onset-rhyme divide holds that (in many dialects) coronal consonants cannot be followed by the diphthong [ju], e.g. *beauty* [bjuɾi] vs. *duty* [duɾi], *[djuɾi], etc. (see, for example, Hammond 1999).

A second traditional argument for this division comes from stress. It is generally thought that only material from the rhyme of a syllable contributes to the weight of a syllable (see, for example, Hayes 1980). Thus, if stress is attracted to 'heavy' syllables in a language only material from the rhyme can contribute to this weight. An example of such a language is Creek (Haas 1977). Accent falls on the final syllable or the second syllable from the right, whichever is an even number of syllables from the left, except that a heavy syllable restarts the count.

Os'ana 'otter'
hitot'iː 'snow'
'icki 'mother'

Evidence for the nucleus–coda division is a little harder to come by, but has also been adduced from stress systems. There are a number of languages where stress is attracted to syllables where only the nucleus contributes to syllable weight. For example, in Khalkha Mongolian (Street 1963), the leftmost long vowel is stressed. If there is no long vowel, the first syllable is stressed.

bosg'uul 'fugitive'
bari'aad 'after holding'
'ali 'which'
xoyərduɡ'aar 'second'
x'ötəlbərə 'leadership'
ɡar'aasaa 'from one's own hand'

Closed syllables are irrelevant to the computation of stress in Khalkha.

While terms like onset, rhyme, nucleus, and coda are still quite widely used, the world of subsyllabic constituency was thrown into question with the publication of Hayes (1989). Hayes argued against these notions and for a *moraic* characterization of the syllable. The basic idea—according to Hayes—is that there are no subsyllabic constituents. Instead, there is only the *mora*. The mora is a unit of syllabic quantity. Light syllables have one mora; heavy syllables have two (or more) moras. If we assume that English syllables have a maximum of two moras, we might represent the English word *trounce* as follows:
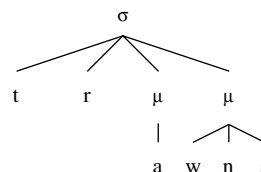
Onset consonants are direct dependents of the syllable node. Elements that we would have formerly put in the

rhyme or nucleus are now docked to mora nodes (abbreviated as Greek mu 'μ'). The number of moras is a function of how weight is assigned in the language in question. The peak of the syllable is generally thought to be always moraic. Long vowels are always bimoraic. Coda consonants may or may not generate their own mora. (There are controversial aspects of the proposal. For example, onset consonants always attach to the σ, or can they dock to the first mora? Can there be moraless syllables? Do moraless coda consonants attach to the final mora or directly to the σ?)

The moraic theory of the syllable accounts directly for quantity–sensitivity in stress systems. If stress is attracted to closed syllables, it is because coda consonants are moraic in the language in question. If stress is attracted to long vowels, but not closed syllables, it is because coda consonants are *not* moraic in the language in question.

Where the moraic theory makes new predictions is its rejection of any sort of onset constituent. Hayes argues that this is required to account for *compensatory lengthening,* the phenomenon whereby a vowel lengthens when a neighboring consonant is deleted. For example, Hayes cites Latin, where an [s] is deleted before a coronal sonorant, triggering lengthening just in case the [s] was a coda. Thus *[kasnus] 'gray' is realized as [kaːnus] while *[snurus] 'daughter-in-law' is realized as [nurus].

Under moraic theory, this has a straightforward analysis. The [s] of *[kasnus] is syllabified as a moraic coda of the first syllable. It is then deleted, leaving the mora behind. The vowel then spreads to the second mora, resulting in a long vowel.
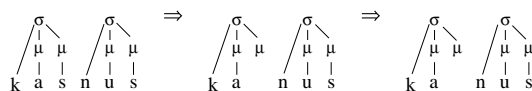
Contrast this with the case of *[snurus], where deletion of the [s] leaves no stranded mora.
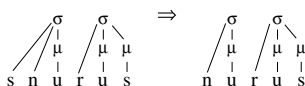
The key generalization about Latin is that only a coda consonant can trigger compensatory lengthening. If an onset consonant deletes, there is no compensatory lengthening. If we had an onset constituent, this would make the *wrong* prediction. A deleted onset might leave some sort of slot for a neighboring vowel to lengthen with. This characterization of syllabic weight accounts for several things. First, it directly describes what can contribute to stress attraction. Second, it accounts for why compensatory lengthening never results from loss of an onset consonant.

With the advent of 'Optimality Theory', subsyllabic constituency has come back into vogue. In their presentation of an optimality–theoretic theory of the syllable, Prince and Smolensky (1983) used the terms onset, peak, and coda. This has led a number of researchers to reintroduce subsyllabic constituency into syllabification theory, although Prince and Smolensky acknowledge that they use those terms only as a convenience. It is an unresolved empirical question as to whether Optimality Theory can be directly reformalized without these subsyllabic constituents.

References

Chomsky, Noam, and Morris Halle. 1968. The sound pattern of English. New York: Harper & Row.

Haas, Mary. 1977. Tonal accent in Creek. Studies in stress and accent, Southern California Occasional Papers in Linguistics 4.

Hammond, Michael. 1999. The phonology of English. Oxford: Oxford University Press.

Hayes, Bruce. 1985. A metrical theory of stress rules. Ph.D. dissertation, Massachusetts Institute of Technology, 1980. New York: Garland.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. Linguistic Inquiry 20

Itô, Junko. 1989. A prosodic theory of epenthesis. Natural Language and Linguistic Theory 7.

Prince, Alan, and Paul Smolensky. 1993. Optimality theory, ms., Brandeis University and University of Colorado.

Street, J.C. 1963. Khalkha structure, Bloomington: Indiana University Publications.

MICHAEL HAMMOND

Syntactic Categories

The words and phrases that make up utterances can be classified as belonging to certain syntactic categories. The criteria for this classification are their function within the clause or phrase, their form, and their distribution, i.e. what other elements they occur with. The term is also used with reference to classes of words and word roots as part of the vocabulary. These two senses are often not distinguished.

A more traditional term that is often used synonymously is 'part of speech'. In traditional grammar, the

following parts of speech were distinguished: verbs, nouns, pronouns, adjectives, adverbs, determiners (articles), prepositions, and conjunctions. This classification goes back to the ancient Greek and Latin grammarians and has been a cornerstone of grammatical description for 2,000 years. It also plays an important role in most modern theories of grammatical structure, although there have been many modifications on the status, the number, and the definition of these categories.

In traditional grammar, the existence of the abovementioned categories was taken for granted. They were axioms in the model of description and served as the background for the definition of other categories and for the analysis of sentences. It was assumed that all languages could be described by using the same set of categories, although the actual forms and functions of nouns, verbs, adjectives, etc., could be different. A similar view is taken within the framework of generative grammar, in which syntactic categories are used to define phrasal categories and remain undefined themselves. It is often assumed that the distinction between nouns and verbs is universal. Furthermore, it is used to characterize other categories; i.e. the four major categories are distinguished with respect to how much they function like a typical noun or verb. In this view, verbs are characterized as verbal and nonnominal, and nouns are nominal and nonverbal. Adjectives are described as having both verbal and nominal properties and prepositions as having neither. This approach allows general statements about categories that include several parts of speech. For example, in some languages, verbs and prepositions, the nonnominal categories, share characteristics that other categories do not have.

A different view was held by the early American structuralists, who challenged traditional grammar in the second quarter of the twentieth century. Instead of assuming universally valid axiomatic categories, they maintained that syntactic categories, or form-classes, should be established on language-particular grounds, that is, on the basis of the formal properties and the distribution of elements of a particular language. A tool for classifying linguistic elements was substitution frames, e.g. 'the ___ is good' for nouns in English. All elements that could fill the slot in such a frame belong to the same distribution class. A careful analysis of the possible substitution frames of words leads to their classification into form-classes. A grammatical description that strictly follows this view comes up with far more classes than the traditional eight parts of speech. Also, these classes were language particular and could not be used for language comparison. For example, English count nouns could be described as a form-class of elements that may appear together with an article and combine with a plural marker. Obviously, another language, which lacks articles or has other means of marking plurality, could

not possess this kind of form-class. If the term noun were used in the description of this language, it would mean something different from what it means in English. For this reason, among others, the traditional part of speech classification was never completely abandoned, and it was revived in generative grammar, which otherwise strove for formal rigor like its predecessor, the structuralistic approach.

Both the traditional parts of speech and the form-classes of structuralist descriptions are primarily concerned with the classification of words. They differ in what aspect of the words they were most interested in.

The object of structuralist form-classes are words as they appear in sentences, units that can be obtained by segmenting utterances into parts and classifying these parts. With this approach, the forms *goes, gone*, and *going* belong to different form-classes because they have different distributions and occupy different slots within an utterance. Traditional grammar, on the other hand, mainly dealt with lexemes, or abstract units of the lexicon that can be realized by one or several word-forms. Here, the lexeme GO, which is realized by the word-forms *go, goes, went, gone*, and *going* is classified only once, namely, as a verb.

The difference between classes of lexemes with their potential forms and functions, and classes of word forms fulfilling a given function in a sentence, has long been ignored. It was obscured by the habit of using the same label for both kinds of categories. For example, the label 'verb' is used both for classes of lexemes like GO and for word-forms used as predicates; in the clause *don't madam me!* (meaning 'don't call me madam'), the word *madam* is a verb in the latter sense, but not in the former.

There are, of course, very obvious links between the two kinds of classes, and these links were particularly strong in the Indo-European languages that had been the model languages for linguistic description for a long time. However, the more linguists explored languages of very different structure, the more it became clear that lexical classes (classes of lexemes) and syntactic word classes have to be established independently of each other. How the two are connected in various languages has been an important topic in the field of language typology in the 1990s. The distinction between the category, or type, of a lexeme and the syntactic category of a word-form has also been integrated into some recent formal theories of syntax.

In a system with a limited number of classes (such as the traditional part-of-speech system), there will always be items that display features of more than one class or that do not fit clearly into one of the classes. This problem can be solved by assuming that classes are not homogeneous but have an internal structure, such that there is a center containing the elements

that meet all criteria typical for the category and a periphery with elements that share only some of these features. This conception of categorization became popular in linguistics in the 1980s, when the concept of prototype was taken over from psychology. Within this view, categories are also allowed to overlap; i.e. words may belong to more than one class.

Another solution lies in categorizing lexemes more than once. Because lexemes have several characteristics that distinguish them and make them similar to other lexemes, several parallel classifications can be established, just as we classify languages according to their historical origin as well as their structural characteristics or their sociolinguistic status. The idea of multiple classification of words has entered linguistic theory only recently and will probably be further elaborated in the future.

The concept of prototype has also been used to solve the old dilemma of universal vs. language-particular categorization. For example, many linguists have been concerned with the question of whether all languages can be said to distinguish nouns and verbs. To tackle this question, a general definition of these categories is needed to ensure that the classes labeled 'noun' and 'verb' are indeed categories of the same kind in different languages. One option is to define the center, or prototype, of the major lexical categories (nouns, verbs, and adjectives) on the grounds of criteria related to meaning and usage. Hence, prototypical nouns are words that designate objects, or time-stable entities, and are used for reference to discourse participants, i.e. what or who is talked about. Prototypical verbs are words that designate actions and are used for the act of predication, and prototypical adjectives designate properties and are used for modification. Within a particular language, one of these categories can be identified if words meeting the general definition are somehow distinguished from other words, i.e. are reasonably considered a lexical class by language-internal criteria such as grammatical inflections and syntactic function. Although the center of each category is determined by the general definition, it is a matter of individual languages as to what further elements the class contains and how classes are delimited. For example, although predicative action words (such as *run*, *push*, *eat*, *sell*) will be verbs in any language that distinguishes such a lexical class, lexemes denoting states or emotions (like *love*, *fear*, *be happy*, *be hungry*) may be verbs in one language and adjectives or nouns in another.

The distinction may be complete, so that e.g. words that are identified as nouns cannot be used in the same syntactic functions and environments as verbs or adjectives. However, we often find that lexemes are not that strictly specified but are used in several functions. In this case, they may be distinguished by the particular word-forms that are possible and necessary in each function. Hence, in many languages, nouns cannot make e.g. tense distinctions, as verbs can, and to have a noun as predicate one often has to use a special device, such as the copula verb in English (*Jane was a girl*, not *Jane girl-ed*).

A language can be said to lack a lexical distinction of parts of speech if any content word can be used in any of the major syntactic constructions and may be combined with the whole range of grammatical markers. Whether such languages exist is still a matter of debate. Certainly, languages differ in how sharp the distinction between lexical classes is drawn, i.e. how important it is for the grammar of the language and whether the elements of a class are more or less specialized for syntactic functions. Latin is an example of a language with clear-cut lexical classes. In English, in contrast, there are many pairs like *to look/a look* or *to move/a move*. Some linguists regard these pairs as two separate lexemes belonging to different classes, whereas others treat them as single multifunctional lexemes.

Whereas the link between a lexical class and a syntactic function is only indirect and often rather weak, matters are different for syntactically determined classes of word-forms. Syntactic theories therefore have been more concerned with the latter, and the term syntactic category is nowadays used in a sense that is only remotely connected to the traditional parts of speech. It covers phrases besides words and also more abstract units that may lack a form, so-called empty categories.

A syntactic category in this sense can be described by a set of features with values specifying the slot in which items belonging to this category can occur. For example, in the clause *she is a linguist*, the category of the word-form *is* will be described by the features 'third person', 'singular', and 'present tense'. As mentioned above, the features 'nominal' and 'verbal' had been derived from the categories noun and verb. Developing this idea further, it becomes possible to characterize a syntactic category entirely by features. In the above example, the syntactic category of *linguist* would thus be the set of the following features: {nominal, nonverbal, singular}. Most contemporary syntactic theories make use of such features but differ in the role attributed to them and their relation to syntactic categories.

Another issue of current debate is the relationship between syntactic categories and sentence structures. In short, the question is which one comes first and should be used to derive the other. This again is connected to the distinction between categories of the lexicon and categories of sentences and how these are related. When sentence structures are derived from word categories, it is supposed that lexemes or word roots are preclassified in the lexicon. For example, the lexicon of English will contain an item *kiss* (noun) and an item *kiss* (verb), the

former defining a noun phrase as in *she gave him <u>a big kiss</u>*, and the latter a verb phrase as in *you may <u>kiss the bride</u>*. According to another position, it is the respective construction that makes *kiss* a noun in the first example and a verb in the latter, whereas in the lexicon, the item *kiss* is not categorized.

References

Bhat, Shankara D.N. 1994. The adjectival category: criteria for differentiation and identification. Amsterdam, Philadelphia: Benjamins.

Croft, William. 1991. Syntactic categories and grammatical relations: the cognitive organization of information. Chicago: University of Chicago Press.

Crystal, David. 1967. English. Lingua 17.

Davis, Henry, and Lisa Matthewson. 1999. On the functional determination of lexical categories. Revue québécoise de linguistique 27(2).

Huddleston, Rodney. 1988. English grammar: an outline. Cambridge: Cambridge University Press.

Lefebvre, Claire, and Lisa Travis (eds.) 1999. A propos des catégories syntaxiques. = Revue québécoise de linguistique 27-2.

Plank, Frans. 1997. Word classes in typology: recommended reading. Linguistic Typology 1.

Radford, Andrew. 1997. Syntactic theory and the structure of English. A minimalist approach. Cambridge: Cambridge University Press.

Sag, Ivan A., and Thomas Wasow. 1999. Syntactic theory. A formal introduction. Stanford: CSLI Publications.

Sasse, Hans-Jürgen. 1993. Syntactic categories and subcategories. Syntax, ed. by J. Jacobs, A. von Stechow, W. Sternefeld, and T. Vennemann. Berlin and New York: de Gruyter.

Vogel, Petra, and Bernhard Comrie (eds.) 2000. Approaches to the typology of word classes. Berlin: Mouton de Gruyter.

NICOLE NAU

See also **Empty Categories; Function Words; Lexicon: Overview; Phrase Structure; Word**

Syntactic Typology

Syntactic typology is one of the major branches of linguistic typology, aiming to classify human languages into different types on the basis of their shared syntactic properties, which are not necessarily the result of common origin or geographical contact. Like other branches of linguistic typology, the ultimate goal of such studies is centered on the provision of (syntactic) universals: those qualities that tend to be common in most, if not all, languages, and enable linguists to make hypotheses in their further studies of languages.

The development of syntactic typology owes a lot to the American linguist Joseph Greenberg, whose interest in typology and language universals led him to propose the theory of 'implicational' universals based on syntactic properties (see below).

There are some criteria based on which languages can be typologized syntactically; some of the most important ones are:

(1) Word order. This criterion is the widest known and the oldest way of classifying the languages syntactically. It was so common in the 1960s that one could easily claim that syntactic typology is mainly known to be represented through word-order studies, i.e. how different lexical items are arranged in various languages, to produce syntactic universals.

Studies have shown that, in a transitive clause which mainly has the elements S, V, and O (which stand for Subject, Verb, and Object, respectively), there are logically six possible combinations: SVO, as in English, French, Hausa, Vietnamese, and Thai; SOV, as in Persian, Japanese, Korean, Amharic, Tibetan, Turkish, and Georgian; VSO, as in Welsh, Tongan, Classical Hebrew, Classical Arabic, and Tagalog; VOS, as in Malagasy, Tzotzil (Mexico), and Cakchiquel (Guatemala); OVS, as in Hixkaryana (Brazil) and Barasano (Colombia); and OSV, as in Xavante (Brazil) and Apurina (Brazil).

These orders are, of course, the 'unmarked' order of elements in languages, an order sometimes being called 'canonical order'. Obviously, in languages with rich inflectional systems and therefore 'free word-order' patterns, the canonical order may be altered and changed.

However, these possible compositions are not equal with respect to their frequency. It has been proved that over 95% of the world's languages use one of the patterns SOV, SVO, and VSO as their basic word order, and object-initial languages are extremely rare and uncommon.

Since not all languages express overt subjects, the relative order of verb and object is often considered to be the most important from the point of view of typology. The main interest in classifying languages in this way lies in the 'implicational' relationships, in that we believe certain word orders make it likely for other constructions to occur (more technically speaking: if

X, then Y). For example, an SVO language, such as English, is likely to have prepositions rather than postpositions (such as *up the tree* rather than **the tree up*), and auxiliaries before main verbs (such as *John may come* rather than **John come may*). Such a language is also likely to have relative clauses (beginning with who/which, etc.) after the noun they refer to, such as *The boy who broke the glass escaped* rather than **The who broke the glass boy escaped*. Similar implications may also be provided for other word orders; for example in a VSO language, we have prepositions rather than postpositions, while in an SOV, postpositions are more likely to appear than prepositions.

The general principle behind these observations appears to be a preference for consistency in the position of the main word (or 'head') in any construction with regard to the items attached to it (or the 'modifiers'). Hence, an SVO language such as English is a 'head first' language and an SOV language such as Persian is a 'head last' language. In this case, the linguist is interested in knowing if, for example, genitives appear before or after the noun heads, or whether articles, adjectives, and numerals come before or after the noun heads in a language with a particular word-order.

Although being a very informative source for linguists, word-order typology has its own flaws because (1) it is only relevant to languages with Subjects and Objects, (2) it only deals with languages with a basic word order, while many languages have mixed word orders, and (3) it is only relevant to basic sentences within a language (declarative, nonpronominal, etc.).

The recent developments in syntactic typology have provided use with some other criteria as well (see Nos. 2–9), but basic word order is always standing out as the most fundamental feature of syntactic typology.

(2) Agreement. Agreement refers to a formal relationship between elements in which a form of one word requires a corresponding form of another. This correspondence usually takes place with respect to case, person, number, and gender. For example, in Arabic, noun and their adjectives must correspond to one another in number , gender, and case. Generally, we can assume three important domains in grammatical agreement: (I) The agreement between subject and verb phrase in terms of person and number, as in *I go* and *she goes*. Some languages, however, represent object–verb agreement, like Swahili, Basque, and Caucasian languages (see *Ergativity* below). (II) Nominal agreement with elements related to noun, such as determiners, adjectives, and appositions. Like the German sentence *Sie sucht einen Jungen, ihren kleinsten Sohn* (she is looking for a boy, her youngest son), where the words *Jungen* (boy) and *Sohn* (son) are both accusative and masculine. (III)

Agreement between subject and predicate, in terms of gender, number, or case; as in *He is a waiter/She is a waitress*. This may also show itself in structures where a word refers back to something mentioned previously, as in *I know Jane quite well, **she** was my secretary for long*.

In this case, a syntactic typologist tries to find out if different verb affixes mark person, number, and gender of subject, object, or other noun phrases, and if such agreement is obligatory.

(3) Universals of negative position. Negation is a syntactic (or more technically, "morphosyntactic") process in which a lexical item denies or inverts the meaning of another lexical item or construction. The lexical item that expresses negation is called negator. The position of a negator as a syntactic element is therefore a criterion for syntactic typology; that is, if the negator comes before or after the verbs, or in prefixing verbs, if it is attached to the whole verb, or is placed before the verbal part.

(4) Case-marking patterns. Case is a way of showing grammatical relationship between certain kinds of words by variations in word structure. Languages make use of different means for marking their cases. In some languages, case may be reflected by the use of affixes. In these languages, we either have prepositions (like English and Arabic) or postpositions (like Turkish) or both (like Estonian). There are, of course, some languages that do not make use of adpositions at all, and make use of variations in word endings, like Latin, Estonian, Russian, and Finnish.

On the other hand, in noninflectional languages (like French and English), where sentence structure or word order encode the syntactic functions of the words, we may associate cases with specific syntactic positions.

(5) Wh-movement. Syntactic typologists are also interested in the position of the question words in sentence, and see if it is initial, final, or other.

(6) Relative clause structures. As an implicational result of word-order studies, the questions concerning the use of relatives before or after noun heads is an issue within the domain of syntactic typology.

(7) Causative constructions. A causative is a grammatical or lexical indication of the causal role of a referent in relation to an event or state expressed by a verb. A causative may be indicated by a verbal affix (*black-black**en**, able-**en**able*), or a component of meaning in the verb, or a special construction, like

George had John killed, in comparison with *The sun solidified the mixture.*

(8) Ergativity. In some languages, the form of the object in a transitive clause is the same as that of the subject of an intransitive clause. Such languages are called 'Ergative' languages, as opposed to languages like English, which are typologically entitled 'nominative-accusative' languages. In studying the ergative languages, the linguist has to check if the subject of a transitive clause has a different form; i.e. in some languages, in intransitive clauses some subjects are treated like transitive subjects, and others like transitive objects. A remarkable number of Caucasian (Georgian, Ubykh, Avar), Austronesian (Tongan), Australian (Dyirbal), and Mayan (Tzeltal) languages are ergative.

(9) Pro-drop languages. Some languages can optionally drop pronouns at the beginning of sentences, as in Persian *man Irâni hastam/Irâni hastam* (I am Iranian/Am Iranian) and one that does not usually do so, such as English. Pro-drop languages seem to behave somewhat differently over a range of constructions from languages that do not drop their pronouns.

References

Aissen, Judith. 1999. Markedness and subject choice in optimality theory. Natural Language and Linguistic Theory 17. 673–711.

Austin, Peter and Joan Bresnan. 1996. Non-configurationality in Australian aboriginal languages. Natural Language and Linguistic Theory 14. 215–68.

Bussman, Hadumod. 1996. Routledge dictionary of language and linguistics. London and New York: Routledge.

Comrie, Bernard. 1995. Language universals and linguistic typology. Chicago: University of Chicago Press.

Legendre, Géraldine 2001. An introduction to optimality theory in syntax. Optimality-theoretic syntax, ed. by Géraldine Legendre, Jane Grimshaw and Sten Vikner, 1–27. Cambridge, MA: MIT Press.

Lyovin, Anatole V. 1997. An introduction to the languages of the world. New York and Oxford: Oxford University Press.

Greenberg, Joseph (ed.) 1963. Some universals of grammar with particular reference to the order of meaningful elements. Universals of human language, 73–113. Cambridge, MA: MIT Press.

Mithun, Marianne. 1987. Is basic order universal? Coherence and grounding in discourse, ed. by S. Tomlin Russell, 281–328. Amsterdam: John Benjamins.

Steele, Susan. 1978. Word order variation: a typological survey. Universals of human language, Vol. 4, ed. by Joseph Greenberg, 585–623. Stanford: Stanford University Press.

BEHROOZ MAHMOODI-BAKHTIARI

Syntax

Human languages are productive systems for relating sound (or more generally gesture, as in signed languages) and meaning. The systematic properties on the sound side are the realm of phonology while those on the meaning side belong to semantics. What is in between is syntax. One of the two most fundamental aspects of syntax is that utterances in a human language are structured. Specifically, a sentence is not merely a string of sounds. Minimally, the sounds must be grouped together into words. But a sentence is not just a string of words either. The words are grouped together into units, called 'constituents'. Consider a simple sentence such as:

(1) The man left.

This is not simply a sequence of three autonomous words. *The* and *man* are closely associated in a way in which *man* and *left* are not. The former pair of words constitute a unit based on the noun *man*, as the following two tests indicate. First, a 'pro-form' (in this case, a pronoun) can substitute for *the man*:

(2) He left

No comparable substitution is possible for the sequence *man left* in (1). Second, an adverb modifying the sentence can easily be inserted between *man* and *left*, as in (3), but not between *the* and *man*, as in (4):

(3) The man obviously left

(4) *The obviously man left [* indicates an ungrammatical sentence, a sequence of words that is not a sentence of the language under investigation]

This pattern of data indicates that the sequence *The man* constitutes a unit while the sequence *man left* does not. There is a division of sentence (1) into two constituents. The first part, the subject, is based on a noun (so is often called a noun phrase (NP)), and the second part, the predicate, on a verb; hence a verb phrase (VP). A common notation for representing this division into units, and the 'names' of the units, is displayed in (5).

(5) [$_S$ [$_{NP}$ The man] [$_{VP}$ left]]

The interpretation of (5) is that the sequence *The man* is a unit of structure, and this constituent is called an

NP; *left* is a unit of structure labeled VP; and the entire sequence of words is a unit of structure called S (Sentence). A more complete structural representation of (1) would also indicate the categories of the individual lexical items *The*, a determiner, *man*, a noun, and *left*, a verb. At this point, the labeled bracketing notation becomes rather cumbersome. A standard alternative more graphically represents the structure in the form of a phrase structure 'tree' as in (6).

(6)

```
            S
          /   \
        NP     VP
       /  \     |
     Det   N    V
      |    |    |
     The  man  left
```
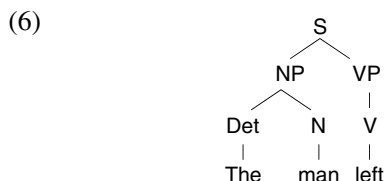
The second fundamental syntactic property of human languages is their infinitude. In every human language, ever longer sentences can be created by embedding one clause inside another, a process called recursion. For example, the sentence in (1) can become the 'direct object' of another verb, as in (7), with tree representation (8).

(7) The woman said the man left

(8)

```
              S
           /     \
         NP       VP
        /  \      /  \
      Det   N    V    S
       |    |    |   /  \
      The woman said NP  VP
                    /  \  |
                  Det  N  V
                   |   |  |
                  the man left
```
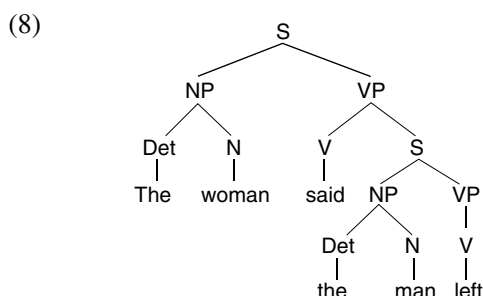
And (7) can, in turn, be inserted into a still larger sentence:

(9) The child believes the woman said the man left

And so on, with no limit. As a consequence of this recursion, there is no longest sentence, and, thus, there are an infinite number of sentences.

The kind of recursion just seen involves 'subordination', placing one unit of structure inside another. Just as general in the languages of the world is coordination — putting constituents together in parallel with a coordinating conjunction like *and:*

(10)

```
               S
            /     \
          NP       VP
        /  |  \     |
      NP  and  NP   V
     /  \     /  \   |
   Det   N  Det   N left
    |    |   |    |
   The  man the woman
```
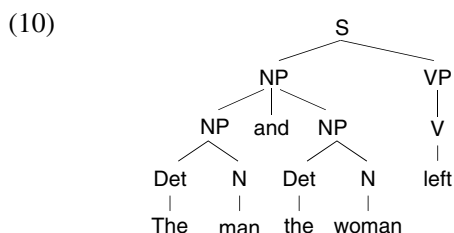
This process too can go on without limit:

(11) The man and the woman and the child left
(12) The man and the woman and the child and the dog left

Another phenomenon that falls in the domain of syntax is a particular kind of ambiguity, 'structural ambiguity'. Human languages are filled with ambiguity. Some instances of ambiguity are *lexical*, involving two different words that happen to be pronounced and/or spelled alike. An example is

(13) Meet me at the bank

where I could intend that we gather after depositing our paychecks or that we get together to go fishing. But many other instances of ambiguity do not depend on homonymous words. Each of the following examples feels like it is two different sentences in one, even though the words of each have constant meaning.

(14) Sue interviewed a singer on the stage
[Was the interview on the stage, or just the singer?]

(15) Susan said John resigned in order to protect his reputation
[Was the resignation or the statement to protect John's reputation?]

These ambiguities can be captured by positing two different syntactic structures for each two-way ambiguous sentence. For (14), the prepositional phrase (PP) *on the stage* can be inside the direct object NP (where it will modify *singer*) or outside that NP, but still inside the VP (where it will modify *interview*). The two structures are in (16) and (17), respectively.

(16)

```
         S
       /   \
     NP     VP
      |    /  \
      N   V    NP
      |   |   /  |  \
     Sue interviewed Det  N   PP
                      |   |    /  \
                      a singer P   NP
                              |   /  \
                             on Det   N
                                 |    |
                                the  stage
```

(17)

```
           S
         /   \
       NP     VP
        |    / | \
        N   V  NP   PP
        |   |  / \   / \
       Sue interviewed Det N P  NP
                        |  |  | / \
                        a singer on Det N
                                     |   |
                                    the stage
```
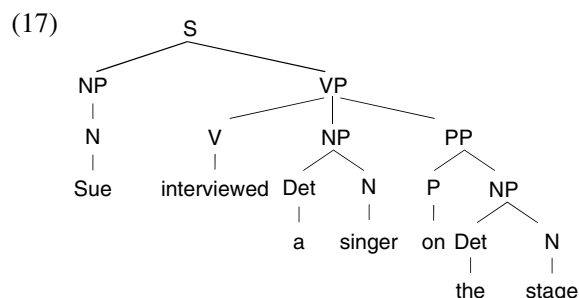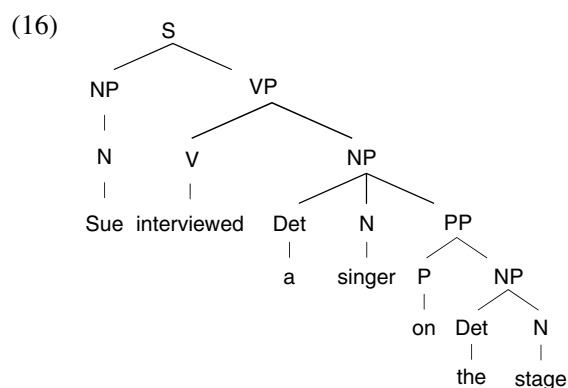
A further striking syntactic property of many human languages is the existence of dependency relations between an overtly realized constituent (a 'filler') and a 'gap' (marked by __), as in interrogative sentences like (18).

(18) Who will Mary hire __

Who, the filler, is pronounced at the beginning of the sentence, but is nonetheless interpreted as the direct object, i.e. in the gap position following the verb *hire*, parallel to (19).

(19) Mary will hire someone

One common way, although not the only way, to represent this kind of dependency is through a 'movement transformation' whereby the interrogative word originates in its understood position, then 'moves' to its superficial position at the beginning of the sentence, leaving a gap in its original position.

Such filler-gap dependencies can span a great distance, as shown in (20), where the edges of each embedded clause are marked with brackets, and __ marks the 'understood' position of the interrogative word.

(20) Who did Mary think [that John said [that Susan visited __]]

Compare:

(21) Mary thinks that John said that Susan visited someone

The properties of long-distance relations of this type ('WH-dependencies', so-called because in English, most interrogative words begin with <u>wh</u>) have been a major concern of syntacticians.

While WH-dependencies can occur across unlimited distances, under some circumstances they are heavily bounded. Constraints on these dependencies (sometimes called 3'island constraints') have been another major topic of investigation. One island constraint involves structures built with coordinating conjunctions. Consider the hypothetical question form of (22) in (23):

(22) Mary will blame [Sue and someone else]
(23) *Who will Mary blame [Sue and __]

This is a perfectly coherent thought, but it cannot be expressed with a WH-dependency. A coordinate structure is an island for such dependencies.

Two other types of dependencies involve an overt item in the subject position and a gap in a 'lower' position. At first glance, the examples in (24) and (25) seem identical to each other, but they turn out to have significantly different properties.

(24) Mary is eager [__ to solve the problem]
(25) Mary is likely [__ to solve the problem]

In both (24) and (25), *Mary* is the understood subject of *solve the problem*. In (24), additionally, the property of 'eagerness' is attributed to Mary, while in (25), 'likeliness' is not. This contrast is made particularly clear in sentences with pleonastic (meaningless) subjects:

(26) *There is eager [__ to be a thunderstorm]
(27) There is likely [__ to be a thunderstorm]

The kind of dependency in (24), where the overt subject is understood in both sentences, is often called 'Control'. The contrasting kind in (25) is sometimes called 'Raising' (a term coming from a movement analysis: *Mary* originates in the lower sentence, where its semantic role is determined, and moves into the higher sentence). Passive sentences display a similar dependency, this time between object position and subject position:

(28) John was chosen __

John is the semantic object of the verb, but is phonologically realized in subject position. A particularly interesting dependency of roughly this sort is illustrated in (29).

(29) Mary is easy [__ to please __]

Here, *Mary* is understood as the object of the embedded sentence. (The subject of the embedded sentence is understood as free in reference.) Syntacticians have long been fascinated by contrasts like these, where, item by item, a set of sentences seem identical, but have very different interpretive properties, as in the following three examples, based on the ones just discussed.

(30) Mary is eager to please [*Mary* is understood subject of both sentences]
(31) Mary is likely to please [*Mary* is understood subject of embedded sentence only]
(32) Mary is easy to please [*Mary* is understood object of embedded sentence only]

In addition to filler-gap dependencies, languages also display dependencies between pairs of overtly realized items. One such dependency involves reflexive expressions, as in (33).

(33) John criticized himself

Typically, a reflexive requires an 'antecedent', and the antecedent must be 'higher' in the tree than the reflexive:

(34) *Himself laughed
(35) *Himself criticized John

Reflexive dependencies, like many other dependencies, have locality restrictions. In English, the antecedent must generally be in the same clause as the

reflexive (although many other languages have long-distance reflexives):

(36) *John said that [Mary criticized himself]
(37) *John said that [himself would do the job]

This is a stricter requirement than we saw for interrogative filler-gap dependencies (as in (20)). On the other hand, in contrast to (23), coordinate structures are not islands for reflexives:

(38) John criticized Mary and himself

One major area of syntactic investigation centers on the characterization of divergent locality constraints for the various kinds of dependencies.

Syntacticians have long been concerned with systematic ways that languages differ in their syntactic patterns. For example, while many languages are like English in having filler-gap WH-dependencies, many others, such as Japanese, realize their interrogative words in their 'understood' positions. In terms of movement analyses, interrogative words in these languages do not move. Another systematic difference among languages involves basic word order. English, along with many other languages, has the verb preceding its direct object, and, parallel to this, English has prepositions (i.e. prepositions precede their objects):

(39) Naomi hired Ken

(40)
```
        . . . . VP
             /     \
           V        NP
           |         |
         hired       N
                     |
                    Ken
```
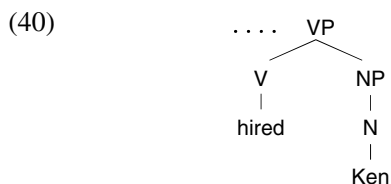
(41) A package arrived from Tokyo

(42)
```
        . . . . PP
             /     \
           V        NP
           |         |
         from        N
                     |
                   Tokyo
```
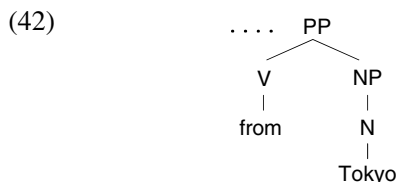
In Japanese, along with many other languages, the structures are exactly the reverse. Verbs follow their direct objects, and the language has *post*positions instead of *pre*positions:

(43)
```
            VP
          /    \
        NP       V
        |        |
        N        N
        |        |
      Ken-o    yatot-ta
```

(44)
```
            PP
          /    \
        NP       P
        |        |
        N       kara
        |
      Tokyo
```
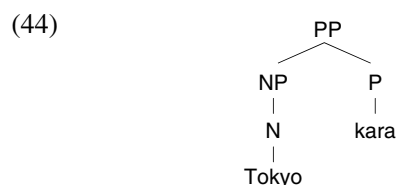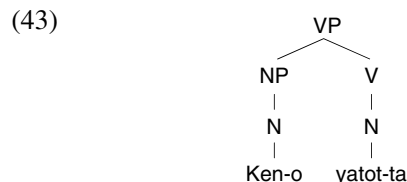
The lexical item that a phrase is based on (V for VP, P for PP, etc.) is commonly called the 'head' of the phrase. The constituent that serves as 'direct object' of the head is the 'complement'. In these terms, the two language types are characterized as head-complement and complement-head.

In sum, syntax is concerned with the structural properties of sentences in languages. Of equal interest are the general, universal properties, such as phrase structure, recursion, dependencies; and the systematically diverging properties, such as relative order of heads and their complements and how dependency relations are realized.

References

Brown, Keith, and Jim Miller. 1991. Syntax: a linguistic introduction to sentence structure. London and New York: Routledge.

Carnie, Andrew. 2002. Syntax: a generative introduction. Oxford: Blackwell Publishers.

Chomsky, Noam. 1988. Language and problems of knowledge. Cambridge, MA: MIT Press.

Green, Georgia, and Jerry Morgan. 2002. Practical guide to syntactic analysis. Stanford: CSLI Publications.

Tallerman, Maggie. 1998. Understanding syntax. New York: Oxford University Press.

HOWARD LASNIK

See Also **Anaphora; Long-Distance Dependency; Phrase Structure**

T

Taboo

Discussions of 'taboo' words must begin with an understanding of the relationship between language and culture. The term originally comes from Polynesian cultures, in which certain objects, actions, and words were considered to exert harmful power on people. Some words were thought to be dangerous. If someone mentioned the name of an ancestor, for instance, negative consequences could result. Thus, speakers of a language had to avoid using certain names or words, just as they had to avoid offending cultural norms and religious beliefs. Which words were considered 'taboo' and to what degree any words were harmful depended on cultural norms.

Societies that recognize supernatural powers and are concerned about offending nonhuman powers consider words to be very powerful. Cultural norms discourage people from using certain words because the very utterance will cause harm either to the speaker or to other members of the society. Melanesian cultures often have a prohibition on mentioning the name of ancestors to avoid any negative repercussions from the spirit world. Since personal names are related to common words, the vocabulary of some languages has undergone a remarkably rapid change. Speakers of such languages find alternative ways of referring to ancestors and find other ways of talking about the objects that are included in the ancestor's name. These avoidance practices may also extend to how a person can talk about or address in-laws.

The discussion of taboo words in the Western world is also related to cultural norms, but not typically to the supernatural world. Certain words may be avoided because they are considered to be a violation of norms of etiquette. The use of these words is unacceptable to certain people in certain situations. As a society shapes the individual, it passes on a form of moral judgment about particular words. In such a society, there is no supernatural judgment or appeal to spiritual powers, it is rather a societal pressure against the use of these words and phrases.

Linguists have taken a neutral and descriptive stance on taboo words. The role of linguistic studies has been to document which words are avoided in what situations. Linguistics makes no moral or religious judgment on the use of taboo forms. Thus, for example, the journal *Maledicta: The International Journal of Verbal Aggression* includes articles listing taboo forms used by various groups and subgroups of a culture.

In English, words related to certain bodily functions (death, defecation, urination, sexual intercourse) and religious beliefs (hell, damnation, curse, swear) are sometimes considered to be 'taboo', at least in particular segments of the population. Those who consider the words to be offensive will often use a synonymous euphemism in place of the taboo word: 'go number two' instead of 'defecate', 'pass on' instead of 'die'. On the other side of the politeness scale are words that are less neutral and more likely to be limited in public settings: 'take a shit' for 'defecate', 'croak' for 'die'.

Words themselves are not 'taboo', 'dirty', or 'profane'. Many of the words currently considered inappropriate in public settings were the neutral, normal term for an object or action in earlier forms of English.

The term 'shit' was not always deemed inappropriate or impolite. In a similar way, many languages of the world still treat bodily functions in a less euphemistic manner.

Since the use of particular words are determined by the social setting, people use or avoid these forms based on their desire to follow or violate the social norms of the particular situation. Violating social norms may be considered appropriate when those norms are thought to be outmoded or prudish. American humorists such as Lennie Bruce, for example, used taboo words in their routines. Young people might use taboo words to show their disregard for the norms of their elders.

The topic of taboo words has been extended to 'cursing' and 'swearing'. The word 'cursing' is tied to a worldview in which bad or evil can be caused by using words. Likewise, 'swearing' has changed from 'taking an oath'. Both words are now used to refer to offensive language. The terms are used as synonyms for 'profanity', which has its roots in the religious notion of the 'profane' vs. the 'sacred'. Such words are also deemed to be 'dirty' as in 'dirty language', 'filthy words', etc. Newspapers and other public media sources typically restrict the use of particular words, even in direct quotations. Likewise, government regulations of broadcast media require certain words to be censored ('bleeped out').

Racial epithets are also considered to be within the topic of language taboo. The use of racial epithets can be seen as a form of aggression against others. The use of names and words that denigrate a racial group is subject to societal norms. In the United States, there are even laws restricting and condemning such language.

Even though portions of society and even government regulations seek to limit the use of taboo forms, it is the individual speaker who determines his or her own use of these forms. In informal settings among co-workers or in dangerous situations (war, natural catastrophe, etc.), people may use taboo forms as a means of expressing their emotions. In more formal situations, in settings that are less homogeneous or in less stressful situations, the use of taboo language is often more limited.

References

Hughes, Geoffrey. 1998. Swearing: a social history of foul language, oaths and profanity in English. London: Penguin.

Montagu, Ashley. 2001. The anatomy of swearing. Philadelphia, PA: University of Pennsylvania Press.

Reinhold, Aman. 1987. The best of Maledicta. Philadelphia: Running Press.

Spears, Richard A. 1998. Forbidden American English: taboo American English. New York: McGraw Hill.

PETER J. SILZER

Tagalog and Philippine Languages

Philippine Languages

Over 150 languages are spoken by the more than 76,500,000 Filipinos who live in an archipelago of around 7,000 islands that stretches over 1,500 kilometers from north to south, and about 800 kilometers from the most western point of Palawan to the most easterly point of Mindanao. Most of the languages are dialectally diverse, with a number constituting extensive dialect chains.

All Philippine languages belong to the Western Malayo-Polynesian group of the Austronesian language family. The archeological record suggests that the earliest Austronesian speakers arrived in the northern Philippines, probably from what is now called Taiwan about 4,500 years ago, at the beginning of the migrations that later took them to the Indo-Malaysian archipelago, and ultimately into the Pacific. Since the time of the first Austronesian habitation of the Philippines, the original language has diversified into a number of fairly clearly defined subgroups (Blust 1991).

Of the 167 Philippine ethnolinguistic groups cited in Grimes (2000), at least ten are presently spoken by more than a million population, and constitute what have been referred to as the 'major' languages of the Philippines. These are Tagalog, Sebuano, Ilokano, Hiligaynon, Bikol, Samar-Leyte, Kapampangan, Pangasinan, Maranao, and Magindanao. The first four of these languages have considerable importance in the Philippines. Tagalog, with some 15,000,000 native speakers, is the native language of Manila and a number of surrounding provinces. It is also the basis of the national language, now known as Filipino. Furthermore, it has become the main language of

movies and comics, and much of the Philippine mass media. It is required to be taught in all the schools in the Philippines, and is rapidly becoming the main second language that people speak throughout the country. Sebuano, Ilokano, and Hiligaynon are widely spoken as regional trade languages. Ilokano is the main language of trade and wider communication spoken throughout northern Luzon. It is also spoken in some areas of southern Mindanao and is the main Philippine language spoken in the United States and other countries to which Filipinos have migrated. Sebuano is used not only in the Visayan area of the Central Philippines, but also in much of southern Mindanao. Hiligaynon is also spoken in some parts of southern Mindanao.

At the other end of the scale, there are scores of languages spoken by relatively small groups of peoples. Many of these languages are spoken in highly isolated areas, such as the remote, mountainous areas of Mindanao, Palawan, and Luzon, as well as on small, isolated islands. Some of the smallest languages, and the ones most in danger of dying out in the very near future are those spoken by some of the approximately 30 surviving Negrito groups, descendants of the original non-Austronesian inhabitants of the Philippines, whose original languages were ultimately replaced by those of the culturally more dominant Austronesians. These include Arta, the language of fewer than a dozen remaining members of a group in Quirino Province, the Atta and Agta groups of Kalinga-Apayao and Cagayan Provinces, some of the Dumagat and Alta groups of Isabela and Quezon Provinces, the Sierra Madre mountain range, and the islands off the eastern coast of Luzon, as well as the many Ayta groups of Bataan and Zambales Provinces. Many of these languages are spoken by fewer than 1,000 people. On a slightly larger scale are the languages spoken in the Cordillera Central, the massive mountain range in the center of northern Luzon. These include closely related varieties of Ifugao, Bontok, Balangaw, Kankanay, and Kalinga, some of which have fewer than 50,000 speakers each. Similarly, the number of speakers of several of the Manobo languages of Mindanao is estimated to be no more than 10,000 to 15,000.

Prior to European settlement in the Philippines beginning with the Spanish in 1521, Philippine languages had already been strongly influenced by contact with traders from outside the country. The Chinese (primarily speakers of Hokkien) had established enclaves in major port areas (beginning around the twelfth century CE). Similarly, traders speaking a variety of Malay probably used in Brunei, Malaysia had firmly established themselves in the Manila area at least 100 years prior to the arrival of the Spanish. Their influence on Tagalog was considerable. But it was the Spanish occupation of the Philippines for over 400 years that had the most impact on the languages of the country. Probably every Philippine language has a large number of words that are ultimately of Spanish origin. There is at least one language in the Philippines that shows a far greater influence from Spanish than does Tagalog. This language, called Chavacano or Zamboangueño, is a creole language spoken mainly in Zamboanga City and surrounding areas.

The other major influence on Philippine languages has come from English as a result of the 50 years that America colonized the Philippines. English is still spoken widely throughout the Philippines, is one of the languages used in education, and continues to influence Philippine languages.

Most Philippine languages have sound systems that are relatively simple. Most have between three and six vowels. Tagalog originally had three distinctive vowels, *i*, *a*, and *u*, but two more vowels, *e* and *o*, have developed as a result of influence from Spanish borrowings. Many languages still retain the same vowel system that has been reconstructed for Proto-Austronesian, with four vowels, *i*, *a*, *u*, and *ə* (Reid 1973). Similarly, most Philippine languages have fewer than 18 consonants. They are nontonal, but some have lexical stress and distinguish vowel length. A number of the languages of northern Luzon, such as Inialoi, exhibit complex morphophonemic variation.

Most Philippine languages have only two or three different kinds of syllables. Words are commonly built using just two syllable types, CV and CVC, where C stands for any consonant and V for any vowel. The disyllabic Tagalog word *bahay* 'house' is typical of the great majority of common nouns in Tagalog and other Philippine languages. Verbs are commonly morphologically highly complex, with various prefixes, infixes, and suffixes providing both syntactic and semantic information. A wide variety of reduplicative processes is also found.

Tagalog

Tagalog is a nonconfigurational language (Kroeger 1993). Its basic sentences are predicate initial, with nominal complements typically following the predicate. Noun phrase word order is not rigid, except that actors typically precede other actants, and full NPs which carry nominative case tend to occur later in the sentence. Since there is no copula verb, Tagalog sentences may have a noun, an adjectival form, or a prepositional phrase as the predicate of a sentence: e.g. *Doktor ang bisita.* 'The visitor is a doctor.' *Maganda ang babae.* 'The woman is beautiful.' *Nasa kusina ang regalo.* 'The gift is in the kitchen.' NPs are typically introduced by one of a small number of short, unstressed words, often

referred to as determiners, that provide information about the case, plurality, and personhood of the following noun, as in Table 1. Most common nouns are not inflected for plurality, but may have a plural word *mga* (/maŋa/) following the determiner to mark plurality: e.g. *Tumakbo ang mga batà* . 'The children ran.' Nominative forms typically mark an NP as specific or definite. They may also mark topicalized NPs and definite NPs functioning as predicates.

Tagalog distinguishes three case-marked sets of personal pronouns, as shown in Table 2. In addition to pronouns that refer to the speaker (first person), directly address the hearer (second person), or refer to a third person, Tagalog like other Philippine languages has distinct forms for first person plural exclusive ('we, not you'), and first person plural inclusive ('we all'). Pronouns with dual reference ('we, two'), while common in many Philippine languages, are used only in rural areas where Tagalog is spoken. An irregular combined form of Genitive 1s and Nominative 2p, *kita*, occurs, e.g. *Iniibig kita.* 'I love you.' Third person pronouns do not distinguish gender. Nominative pronouns also function as predicates: e.g. *Ikaw ang anak ni Pedro.* 'You are the child of Pedro.' Genitive forms express not only possessors that follow their noun heads, but also actor participants in a clause: e.g. *Ibinigay nila ang libro sa titser nila.* 'They gave the book to their teacher.' Pronominal forms, as well as clitic adverbs such as *na* 'already', *pa* 'yet', and *din* 'also', occur as second position clitics attaching to the first constituent within the clause, such as a negative auxiliary, e.g. *Hindi ko pa rin binili ang libro.* 'I still haven't bought the book yet', or a fronted adverb, e.g. *Bukas na siya papasok.* 'He will enter tomorrow already.'

TABLE 1   Tagalog Determiners

	Nominative	Genitive	Oblique	Locative
Common	*ang*	*ng* /naŋ/	*ng* /naŋ/	*sa*
Personal				
Singular	*si*	*ni*	—	*kay*
Plural	*sina*	*nina*	—	*kina*

TABLE 2   Tagalog Personal Pronouns

	Nominative	Genitive	Locative
Singular			
1st person	*ako*	*ko*	*akin*
1st pers. dual	*kata*	*nita*	*kanita*
2nd person	*ka / ikaw*	*mo*	*iyo*
3rd person	*siya*	*niya*	*kaniya*
Plural			
1st pers. excl.	*kami*	*namin*	*amin*
1st pers. incl.	*tayo*	*natin*	*atin*
2nd person	*kayo*	*ninyo*	*inyo*
3rd person	*sila*	*nila*	*kanila*

Relative clauses follow their head noun and are linked to it by a 'ligature' *na* (following consonants) or *-ng* (following vowels). Their structure follows the so-called gap strategy, with the gap corresponding to the nominative NP of the relative clause, e.g. *ang babaeng pumasok* 'the woman who entered', *ang bahay **na** bibilhin ko* 'the house that I will buy'. A ligature also links a main verb and its complement, e.g. *Huwag kang umiyak.* 'Don't cry.'

Tagalog is typical of other Philippine languages that have traditionally been considered to be typologically different from other languages in that they have been assumed to have a unique type of grammatical system, known as the 'focus system'. This is a system characterized by the use of verbal affixes to indicate the thematic role of the NP bearing the nominative case in a sentence. The term 'focus system' was first introduced to describe the languages of the Philippines. Subsequently, other Austronesian languages exhibiting a similar type of grammatical system (such as many of the languages in Taiwan, Sabah, northern Sarawak, and northern Sulawesi, as well as Malagasy, Palauan, and Chamorro) have been described as having a 'Philippine-type' syntax.

Basic verbal clauses in Tagalog have one of two basic structures. 'Actor focus' verbs carry one of a set of affixes on the verb that indicate that the actor is expressed by the Nominative case. These may be monadic, expecting only one NP, e.g. with infix *-um-*: *Pumasok si Nila.* 'Nila entered.', or dyadic, in which case an additional NP expressing a theme argument occurs, expressed by the oblique case when indefinite, or a partitive, definite NP, e.g. ***Bum**ili sila ng mangga.* 'They bought mangoes'; ***Um**inom noon ang mga batà.* 'The children drank some of that.' Such sentences are syntactically intransitive. 'Nonactor focus' verbs carry one of a different set of affixes on the verb that indicates whether some participant other than the actor carries Nominative case. In these sentences, the actor carries Genitive case. All are syntactically transitive: e.g. Goal focus, with *–in*: *Bibilh**in** nina Juan ang mga mangga bukas.* 'Juan (and companions) will buy the mangoes tomorrow.' Location focus, with *–an*: *Bibilh**an** nila ng mangga ang mga batà.* 'They will buy mangoes from the children.' Beneficiary focus, with *i-*: ***I**bibili nila ng mangga ang batà.* 'They will buy the child a mango.' Instrument focus, with *ipang-*: ***Ipam**bibili nila ng mangga ang pera nila.* 'They will buy mangoes with their money.'

In addition to the affixes that mark focus or voice, verbs also carry a wide range of affixes marking temporal aspect (perfective, imperfective, and contemplated), and mood (volitional vs. nonvolitional). They may also be derived with a causative affix, introducing an additional actant (a causer) into the clause.

The grammatical system of Philippine-type languages has been a topic of considerable controversy in linguistic analysis. Hardly any of the statements made in this article have gone unchallenged in recent times. Various issues discussed in the literature are the following: Do Philippine languages have a 'subject', and if they do, which NP is it? Are Philippine languages accusative, ergative, split-ergative, or some other type? Do Philippine languages have a true passive construction? Do they have an antipassive construction? Do the 'focus affixes' constitute inflectional voice morphology on the verb, or are they derivational 'applicative' affixes?

References

Blust, Robert A. 1991. The greater Central Philippine hypothesis. Oceanic Linguistics 30. 73–129.

———. 2002. 'Notes on the history of focus' in Austronesian languages. The history and typology of Western Austronesian voice systems, ed. by Fay Wouk and Malcolm Ross. Canberra: Pacific Linguistics.

De Guzman, Videa P. 1998. Ergative analysis for Philippine languages: an analysis. Studies in Austronesian linguistics, ed. by Richard McGinn. Athens, Ohio: Ohio University Center for International Studies-Center for Southeast Asia Studies.

Grimes, Barbara F. (ed.) and Joseph E. Grimes (consulting editor) 2000. Ethnologue: languages of the world, 14th edition. Dallas, Texas: Summer Institute of Linguistics.

Himmelmann, Nikolaus P. 1991. The Philippine challenge to universal grammar. Arbeitspapier Nr. 15 (Neue Folge). Köln: Institut für Sprachwissenschaft, Universität zu Köln.

Kroeger, Paul. 1993. Phrase structure and grammatical relations in Tagalog. Dissertations in Linguistics. Stanford: Center for the Study of Language and Information.

Liao, Hsiu-chuan. 2004. Transitivity and ergativity in Formosan and Philippine languages. Ph.D. Dissertation. Department of Linguistics, University of Hawai'i.

Mithun, Marianne. 1994. The implications of ergativity for a Philippine voice system. Voice: form and function, ed. by Barbara Fox and Paul J. Hopper. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Quakenbush, J. Stephen. 2003. Philippine linguistics from an SIL perspective: trends and prospects. Philippine Journal of Linguistics 34. 1–28.

Reid, Lawrence A. 1973. Diachronic typology of Philippine vowel systems. Current trends in linguistics, Vol. 11: diachronic, areal, and typological systems, ed. by Thomas A. Sebeok. The Hague: Mouton and Co.

Reid, Lawrence A., and Hsiu-chuan Liao. 2004. A brief syntactic typology of Philippine languages. Special issue of Language and Linguistics 5(2) (Issues on Austronesian Linguistics), ed. by Elizabeth Zeitoun. Taipei: Academia Sinica.

Schachter, Paul, and Fe T. Otanes. 1972. Tagalog reference grammar. Berkeley, Los Angeles, London: University of California Press.

LAWRENCE A. REID

Tamil

Tamil is a language of the Dravidian family that originated in southern India and spread in prehistoric times to Sri Lanka. In the nineteenth century, the British colonial empire enabled Tamils from India to seek work as laborers in Malaysia, Singapore, Sri Lanka, Burma, southern Africa, Fiji, Reunion, Mauritius, and the West Indies. Many well-educated Sri Lanka Tamils found higher status jobs in Malaysia and Singapore. More recently, Tamils have looked to the Middle East, the United Kingdom, North America, and Australia for employment. In addition, thousands of Sri Lanka Tamils seeking asylum from the troubles in their homeland have fled to India, Western Europe, North America, and Australia. The distribution of the over 63 million native speakers of Tamil in various countries is roughly as follows: India, 58.6 million, primarily in Tamil Nadu and neighboring states; Sri Lanka, 3 million; Malaysia, 1.3 million of 1.5 million ethnic Tamils; South Africa, 250,000; Canada, 160,000; Singapore, 120,000 of 191,000 ethnic Tamils; Reunion, 120,000, a figure that may include second-language speakers; United Kingdom, 100,000; France, 60,000; Germany, 50,000; Switzerland, 40,000; United States, 35,000; Australia, 30,000; Italy, 25,000; Mauritius, 22,000 of a total of 100,000 ethnic Tamils; Netherlands, 20,000; Norway, 8,000; Denmark, 7,000; Fiji 6,600; New Zealand, 3,000; Sweden, 2,000. In all communities outside India and Sri Lanka, younger generations are no longer learning Tamil, so that in some cases these figures exaggerate the number of actual speakers.

Tamil exhibits 'diglossia': a formal or 'High' variety is used in public speaking and most writing, while colloquial or 'Low' varieties are used in face-to-face interactions and in writing intended to convey an informal chatty or 'hip' style. Differences between formal and colloquial varieties are found in pronunciation, grammar, and vocabulary. Colloquial Tamil has much more variation according to region and social class than formal Tamil does: the greatest differences are between Sri Lanka and India, with different subvarieties within

each of these. Differences in the usage of different social groups are also found, as, for example, between Brahmin and non-Brahmin varieties in India. Some examples are given Table 1 comparing formal Tamil with the principal colloquial varieties of India and Sri Lanka.

In vocabulary that is not borrowed from unrelated languages like Sanskrit or English, formal Tamil has the consonants / p t t̪ t̤ c k m n ŋ ɲ r ɻ l ɭ ʋ j / and the vowels / i e a o u /. All consonants except /r/ and /ɻ/ are found single (short) or geminate (long), and all vowels may be short or long. The consonants and vowels are essentially the same as those believed to have been used in Proto-Dravidian, the common ancestral language of the entire Dravidian family. Unlike the other Dravidian languages, Tamil has not augmented the single series of voiceless stops /p t t̪ t̤ c k / with voiced or aspirated ones, although the voiceless stops have variants that include voiced stops such as [b, d] and continuants; and for many Indian speakers, a series of voiced stops occurs in loanwords such as *bassu* 'bus' and *daakam* 'thirst'. Other sounds that occur only in loanwords are / h /, / ʃ /, and / æ /. Colloquial Tamil has generally lost /ɻ/, replacing it with / ɭ / or / j /; Indian varieties have also lost / t̪ /, but Sri Lanka Tamil preserves it as / r / or / t̪ /. Two characteristics of the pronunciation of colloquial Indian Tamil distinguish it from Sri Lankan varieties: the insertion of / j / before / i / or / e / and of / ʋ / before / u / or / o / when these vowels begin a word; and nasalized vowels that result from the loss of / m / or / n / at the end of a word, as in [pɐjjẽ] from *pajjan* 'boy.' Sri Lanka Tamil, for its part, is characterized by the retracted pronunciation of / i / and / e / before certain consonants, as in [kɜɜ kkɾɐn] from *keekkɾan* 'I ask', and the fronted pronunciation of / a / before alveolar consonants, as in [pælli] from *pallu* 'tooth'.

Tamil has been written since the third century BCE. Like many South Asian literary languages, it has its own

script, which descends ultimately from a variant of the Brahmi script used during the reign of Emperor Ashoka (273–232 BCE). The writing system is well adapted to represent the pronunciation of formal Tamil, except that a few symbols are superfluous, such as those that represent the combinations /aj/ and /aʋ/ as though they were diphthongs, and the symbols for [n̪] and [ŋ], which are just variants of /n/. A second script, the Grantha script, was developed in the fifth century CE to write Sanskrit in South India. Grantha symbols for /dʒ/, /ṣ/, /s/, /h/, and /kṣ/ have been adopted into the Tamil script to represent sounds occurring in loanwords.

A Tamil word consists of a root plus a series of suffixes. Word order is strictly 'head-final': that is to say, adjectives and relative clauses precede the nouns they modify; adverbs, direct objects, and indirect objects precede verbs; auxiliaries follow main verbs; and postpositions are used rather than prepositions. Formal Tamil is more conservative and exhibits characteristics of an earlier stage of the spoken language. Thus, some of the suffixes of colloquial Tamil have developed from earlier postpositions, auxiliaries, etc., which still appear as separate words in formal Tamil. Examples can be seen in the last three examples in the table above.

Case suffixes and postpositions are used to indicate the grammatical functions of nouns and pronouns. In head-final languages like Tamil, the distinction between postpositions and case endings is sometimes fuzzy; indeed, as mentioned above, the latter can develop from the former. To illustrate this point further, traditional grammars of formal Tamil recognize eight cases: nominative (subject), accusative (direct object), instrumental (by using …), associative (together with …), dative (to …), ablative (from …), genitive (of …), and locative (at …), but the ablative is not a case in formal Tamil: rather, it is a combination of the locative case and the postposition *iruntu,* as seen in the example 'from the house' above. In colloquial varieties, however, the postposition and locative have fused into a single unit and can be considered a separate case form.

Pronoun categories include the expected first, second, and third person forms ('I/you/he'). In the first-person plural ('we'), there is an additional distinction (lacking in colloquial Sri Lanka Tamil) between 'inclusive' and 'exclusive' (*naam* 'we, including you' vs. *naankaḷ* 'we, not including you'). The second person distinguishes between the familiar singular *nii* and the plural or polite singular form *niinkaḷ;* a singular polite form, *niir,* is also less commonly found. The third-person pronouns are the most numerous and show the greatest variation from one variety of Tamil to another. First, there are three parallel sets of 'deictic' (pointing) pronouns, which are distinguished by their initial

TABLE 1 Comparison of Formal Tamil with Colloquial Forms of India and Sri Lanka

	Common Indian Colloquial	Sri Lanka Colloquial (Jaffna)	Formal Tamil
[I] gave	kuʈutteen	kuʈuttanaan	koʈutteen
[I] gave (perfective)	kuʈuttiʈʈeen	kuʈuttiʈʈan	koʈuttu ʋiʈʈeen
[you] ask	keekkriinka	keekkrijaḷ	keeʈkiʈiirkaḷ
very	rompa	miccam	mikaʋum
boy	pajjan	poɖijan	pajjan
from the house	ʋiiʈʈulerntu	ʋiiʈʈulerntu	ʋiiʈʈil iruntu
must go	pookaɳum	pooka ooɳum / pooka ʋeeɳum	pooka ʋeeɳum
didn't come	ʋarale	ʋareelle	ʋara ʋillai

vowel; for example, *itu* 'it, this one', *atu* 'it, that one', *etu* 'which one'. These three sets can be called 'proximal' (*i-*), 'distal' (*a-*), and 'interrogative' (*e-*). In formal Tamil, each of these sets has six members; illustrating with the distal set, these are: *avar* 'he or she, that (human, respectful) person', *avan* 'he, that (human, nonrespectful, male) person,' *aval* 'she, that (human, non-respectful, female) person', *atu* "it, that (non-human) one", *avarkal* 'they; those (human) people', and *avajkal* 'they, those (non-human) ones'. In addition to these deictic pronouns, two further interrogative pronouns are found: *enna* 'what' and *jaar* 'who'. Many other deictic words also have proximal, distal, and interrogative forms, for example *ippootu* 'now', *appootu* 'then', *eppootu* 'when'; *ittanaj* 'this many', *attanaj* 'that many', *ettanaj* 'how many'; and *ippati* 'thus, like this', *appati* 'thus, like that,' *eppati* 'how'. Finally, the Jaffna dialect of Sri Lanka Tamil has a fourth set of addressee-proximal deictics in *u-*. For example, Jaffna Tamil distinguishes between *itu* 'it, this one near me/us', *utu* 'it, that one near you', *atu* 'it, that one (not near you or me/us)', and *etu* 'which one'.

Verbs do much of the grammatical work in a Tamil sentence, and as a result there is a rich set of verb forms; a selection of these is given in Table 2. For regular verbs, the verbal root serves as the base of the imperative. The infinitive is also closely related to the root form of the verb. Three tenses (present, past, and future) are found, and these take additional suffixes to agree with the subject. The infinitive and tense stems also serve as the base on which other forms are built; finally the infinitive and the past participle combine with a variety of auxiliary verbs. In colloquial Tamil, some of these auxiliaries are reduced to suffixes, as seen in the example '[I] gave (perfective)' above. Colloquial Tamil has also developed new formations not found in formal Tamil; an example is the past formation of Jaffna Tamil seen above in the example '[I] gave'.

It can be seen from the examples that the human vs. nonhuman distinction found in the pronouns has parallels in the verbal system; for example, the human present and future, built on the corresponding tense stems, are structurally different from the nonhuman present and future, built on the infinitive stem. Less apparent is the fact that the positive and negative systems of Tamil are somewhat independent of one another. In most languages, each positive form has a corresponding negative form with a similar structure. But in Tamil, this is not the case: the distinctions of tense (present, past, future) and aspect (progressive, perfect, perfective) found in the positive are replaced in the negative by just the simple negative, the future human negative, and the future nonhuman negative; of these three, only the future nonhuman negative has

TABLE 2   Selected Forms for the Verb koṭu 'give' (Formal Tamil)

Imperative	
koṭu	imperative 'give'
Infinitive	
koṭu-kk-a	infinitive 'to give'
Tense forms	
koṭu-kkiṯ-aan	rational present 'he gives/is giving'
koṭu-tt-aan	past 'he gave'
koṭu-pp-aan	rational future (potential/habitual) 'he will give; he gives; he used to give'
Forms built on infinitive stem	
koṭu-kk-alaam	permissive/speculative 'may/could give'
koṭu-kk-aṭṭum	Optative 'let [him] give'
koṭu-kk-atu	nonrational present 'it gives/is giving'
koṭu-kk-um	nonrational future (potential/habitual) 'it will give; it gives; it used to give'
koṭu-kk-aatu	nonrational future (potential/habitual) negative 'it won't give; it doesn't give; it used not to give'
Forms built on present stem	
koṭu-kkiṯ-a	pres. verbal adjective '[who] gives'
koṭu-kkiṯ-atu	pres. verbal noun 'giving'
Forms built on past stem	
koṭu-tt-a	past verbal adjective '[who] gave'
koṭu-tt-aal	conditional 'if [he] gave/gives'
koṭu-tt-u	past participle 'having given'
Forms built on future stem	
koṭu-pp-atu	future verbal noun 'giving'
Infinitive + auxiliary	
koṭu-kk-a muṭij-um	habilitative 'can give'
koṭu-kk-a ʋeeṇṭ-um	debitive 'should/must give'
koṭu-kk-a ʋillaj	simple negative 'didn't give/isn't giving/won't give'
koṭu-kk-a maaṭṭ-aan	future rational neg. (potential/habitual; volitive) 'he won't give; he doesn't give; he used not to give'
koṭu-kk-a ppaṭ-utu	passive 'it is given'
koṭu-kk-a ʋaj-kkiṯ -aan	Causative 'he makes [someone] give'
Past participle + auxiliary	
koṭu-tt-u kko-ṇṭ- iru-kkiṯ-aan	pres. progressive 'he is giving'
koṭu-tt- iru-kkiṯ-aan	pres. perfect 'he has given'
koṭu-tt-u ʋi□ṭu-kiṯ-aan	pres. perfective 'he (definitely/completely) gives/is giving'

any structural similarity to the positive form it negates. Moreover, the human future positive is not completely equivalent semantically to the human future negative, which carries connotations of intention.

Among the constructions typical of Tamil and other Dravidian languages are the conjunctive participle, the relative clause, and the quotative. In the conjunctive participle construction, found throughout the South Asian linguistic area, all but the final verb in a sequence of verbs are represented as past participles (emphasized in the following examples): *dʒejalalitaaʋiṉ ʋiiṭṭukku **ppooj** aʋaraj cantittaar* 'He

went to Jeyalalitha's house and met her' (literally 'Jeyalalitha's house **having-gone,** her met'); *tankaɭuʈaja paʈaja dʒanapatijaj aluvalaka kaʈʈattukk uɭ iruntu **taɭɭi** taratara-enʈu **iʈuttu ppooj** naʈucaalajil tuukkil pooʈʈaarkaɭ* "They pushed their own former president out from inside his office building, went dragging him along the ground and hanged him in the middle of the road" (literally "their-own former president office building inside from **having-pushed** along-the-ground **having-dragged having-gone** in-the-middle-of-the-road in-a-noose put").

Relative clauses in Tamil and other Dravidian languages are formed without relative pronouns, by using a verbal adjective (emphasized in the following examples); the relative clause precedes the noun it modifies, and the verbal adjective is the last item in the relative clause: *[naan **koʈutta**] pputtakam...* "The book that I gave…" (literally "I **having-given** book"); *inta aarajccijin muʈivil [ve ɭijaana] viʂajam [nammaj ellaam niccajam atirccikk u ɭɭaakkukiʈa] maatiri taan irukkiʈatu* "The information which has been released in the results of this research is of the very sort that will certainly subject us all to shock" (literally "this research's results-in **having-been-released** information us all certainly shock-to **subjecting** type EMPHATIC is").

Finally, the quotative construction marks the complement of an overt or understood verb of saying, thinking, hearing, etc. with a QUOTATIVE complementizer derived from the verb *en* 'say'. As can be seen in the following examples, Tamil makes no distinction between direct and indirect discourse. *'atil enna viceeʂam enʈu' avacarappaʈa veeɳʈaam* 'Don't be in a hurry [to say] "What's special about that?"' (literally "'that-in what special" QUOTATIVE hurry don't'); *'anta maaʈʈaj naan vaanki-kkoɭkiʈeen' enʈu mookan keeʈʈu-kkoɳʈ-um kuuʈa, 'muʈijaatu' enʈu raaman maʈutt-u viʈʈaar* 'Even when Mohan offered to buy the cow, Raman refused [saying] "[you] cannot"' (literally "'that cow I will-buy" QUOTATIVE Mohan having-asked-even, "cannot" QUOTATIVE Raman refused').

References

Andronov, Mikhail Sergeevich. 1989. A grammar of modern and classical Tamil. Madras: New Century Book House.

Annamalai, E., and Sanford B. Steever. 1998. Modern Tamil. The Dravidian languages, ed. by Sanford B. Steever. London, New York: Routledge.

Asher, R.E. 1985. Tamil, London, Sydney, Dover, NH: Croom Helm.

Britto, Francis. 1986. Diglossia: a study of the theory with application to Tamil. Washington, DC: Georgetown University Press.

Gair, James W., S. Suseendirarajah, and W.S. Karunatilaka. 1978. An introduction to spoken [Sri Lanka] Tamil. Kelaniya: External Services Agency, University of Sri Lanka.

Hart, Kausalya. c. 1999. Tamil for beginners. Berkeley: International and Area Studies, University of California at Berkeley.

Lehmann, Thomas. 1989. A grammar of modern Tamil. Pondicherry: Pondicherry Institute of Language and Culture.

Schiffman, Harold. 1999. A reference grammar of spoken Tamil. Cambridge: Cambridge University Press.

Suseendirarajah, S. 1993. Jaffna Tamil. Thirunelvely, Sri Lanka: University of Jaffna.

Zvelebil, Kamil. 1973. The smile of Murugan—on Tamil literature of South India. Leiden: E.J. Brill.

Zvelebil, Kamil, Yurii Glasov, and Mikhail Andronov. 1967. Introduction to the historical grammar of the Tamil language, transl. by L. Navrozov. Moscow: Nauka.

Web Resources

Grimes, Barbara F. (ed.) 2000. Ethnologue: languages of the World. SIL International,<http://www.sil.org/ethnologue/>

Tamil Electronic Library <http://www.geocities.com/Athens/5180/index.html>

Web Assisted Learning and Teaching of Tamil, University of Pennsylvania Language Centre, <http://ccat.sas.upenn.edu/plc/tamilweb/>

IAN R. SMITH

Tarascan

Tarascan is the name of a language (and its indigenous speakers) spoken in the mountainous north central region of the state of Michoacan, Mexico. Tarascan speakers call themselves and their language *Purépecha,* variously written as Purhépecha, P'urepecha, P'orepecha, Phurhépecha, and sometimes shortened to P'oré or P'uré simply to denote things Tarascan. Many consider *Purépecha* preferable to *Tarascan,* an anglicized version of the Spanish word *tarasco.* The generally accepted etymology of 'tarasco' derives from stories of the first contact with Europeans, who through various means acquired for themselves

daughters of the Tarascan nobles and princes. *Tarháskua* means 'son-in-law', 'daughter-in-law', or 'father-in-law' in the indigenous language, all relationships which, in effect, the Spaniards acquired with the Tarascans. The repeated, and uninformed, use of the word is said to have spread among the Spaniards as a reference to the Indians. The etymology of the word 'purépecha' has two basic strands. Some have traced its meaning to refer to migration, travel, and alliances while other scholars take the word to mean 'common people' *(plebeyos)*.

The geographic range of spoken Tarascan is today limited to the state of Michoacán. Historically, however, the Tarascan empire (known to the Spaniards as the 'Western Empire') was much larger. Prior to European contact, the Tarascans are thought to have exerted their political and cultural influence as far north as the present-day state of Durango. They maintained a militarized border against the Aztecs, established as far east as Guanajuato and the Valley of Toluca, and to the west their control reached the Pacific Ocean. Tarascan presence is thus estimated to have covered an area nearly equal in size to the empire of the Aztecs, extending over virtually all of present-day Michoacán and with settlements or influence in portions of the states of Guerrero, México, Jalisco, Guanajuato, Querétaro, Colima, Nayarit, Sinaloa, and Durango.

Through the centuries the extent of spoken Tarascan has shrunk significantly, to the point where it is now found in a well-defined area comprising about 3,500 square miles in north central and northwestern Michoacán only. Researchers commonly describe four distinct zones of Tarascan speech: the Zacapu Marsh region, the Lake Pátzcuaro region, including the town bearing that name and the villages surrounding the lake, an area called La Cañada de Chilchota (Los Once Pueblos), once containing 11 towns, now nine, to the north and west of the Lake, and the mountainous region due west of Lake Pátzcuaro called the Tarascan Sierra (also known as *La Meseta Tarasca).* The Sierra hosts the largest percentage of Tarascan speakers (an estimated 60% as opposed to 19% in the next largest zone, the Lake Region). Tarascan speakers are distributed in 95 of Michoacán's 113 *municipios.* Researchers have observed that it is commonly thought throughout the region that 'real' Tarascan is spoken in the Sierra.

The number and location of Tarascan speakers has historically been difficult to assess, given the remote nature of many of the villages where the language is spoken. Through much of the middle part of the twentieth century, the numbers were variously pegged at 50,000 to 80,000 speakers. More recently, some linguistic survey counts have put the numbers as high as 100,000 and 120,000. The Mexican National Indigenous Institute *(*Instituto Nacional Indigenista*, or INI) in its most recent count puts the number at just under 100,000 speakers in the state of Michoacán itself. Linguistic surveys have also revealed differing degrees of Tarascan-Spanish bilingualism throughout the region, the smaller, less hispanicized towns, of course, with higher rates of Tarascan use. Approximately 7,217 (8% of the Tarascan-speaking population of the state) are monolingual. Despite the historical decline in numbers of speakers, Mexican census data since 1970 have revealed an average annual increase of 2.6% in the number of speakers, compared to an average annual growth rate of 2.1% for the state's population as a whole. This increase in the number of speakers is owed in part to the work of the Purhépecha Language Academy (La Academia de la Lengua Purhé) and the Center for Investigations of P'urhépecha Culture (Centro de Investigaciones de la Cultura P'urhépecha) (Argueta 2003). The spoken language is characterized by clear dialect differences, some regional (Sierra, Cañada, Lake, as mentioned above), but others in much closer proximity, from village to village (Friedrich 1975).

Interest in the ethnography and language of the Tarascan people dates to the first half of the sixteenth century, shortly after contact with the Europeans. The Franciscans accompanying the Spanish showed immediate interest in the ethnographic and linguistic characteristics of the Tarascans, documenting their findings in the *Relación de Michoacán,* probably begun in 1539 (finished and presented to the Viceroy in 1541). The first systematic linguistic works on Tarascan were undertaken by the Franciscans around the middle of the sixteenth century, and at least three dictionaries and two grammars of the language were produced in the 1600s. The earliest of these works, titled *Vocabulario en la lengua de Michuacan,* was brought to fruition in 1559 by Fr. Maturino Gilberti and subsequently reedited in 1898 by Nicolás León, in 1901 by Antonio Peñafiel, and yet again in 1962 by Ernesto Ramos Meza.

Despite more than four centuries of linguistic analysis, the origins of both the Tarascan language and its people remain puzzling. It is certain that the Tarascan language bears no linguistic relationship with any of the original languages spoken in Mexico. Scholars generally agree that its genetic relationships are difficult, if not impossible, to define for any other languages as well. Attempts have been made on various fronts to trace the origins of the language and its speakers. Linguistic and material cultural evidence is cited to support the idea of a historic connection between the Tarascan language

and the Pacific Coast of South America, most notably with Quechua of Peru, but observed phonetic similarities with indigenous languages of Peru may be merely coincidental. Recent evidence is interpreted by some to indicate linguistic, material culture, and even physiological links to Central American groups. Other scholars do not believe the evidence points unequivocally to a southern provenance and remain open to a northern or northwestern origin. Swadesh (1969) determined Tarascan to be an autonomous linguistic family, a finding with which others concur.

The phonological characteristics of the Tarascan language are relatively clear-cut.

6 vowels:

i	u	e
a	o	

ï (predictably following the consonants s, ts, and ts')

16 Consonants:

P	t	ts	Ch	k
b	d	g	S	J
ill	n	r	X	

rh (a retroflexed 'r')

Five of the consonants take phonologically distinctive aspiration:

p'	t'	ts'	ch'	k'

Despite its straightforward appearance, this phonological pattern is subject to a complex set of morphophonemic rules, which include vowel lengthening, stop epenthesis, nasal assimilation, stop assimilation, retroflexion, vowel deletion, and final devoicing. Lengthy words comprised of many suffixes and multiple roots are a regular feature of the language. Verbs are inflected for person, number, tense, as well as mode and interrogation. Formality is also marked on the verbs (as well as on pronouns). Ordered suffixation characterizes the verbal structure, such that verbs may develop extensive forms. For example:

Unandontskuarhinsïndiksï *(from* uandontskuarhini 'to converse, to chat')
Syllable patterns follow eight types:

V	VV	VC	CV	CVC
CVV	CCV	CCVV		

Word order is generally Subject–Object–Verb.

A sample of written Tarascan (taken from the popular folk song *Dalia Tsïtsïki*) is as follows:

Ay dalia tsïtsïki emenda anapu, ka shankar t'u sesarhati tsïpajka; ka aimindu tumbiecha ishï jarashtia su sha aru komjarhani erontashïn. Isï sha jaueru iamindu tumbiecha nochkan ji tiempushin isï tsïperan, no *para ji trabajimbo ekarha askia.*

The strength of the Tarascan language (as measured by the language surveys of the Mexican Census data) is encouraging. Only speakers over the age of 5 are counted for the language census. Government efforts in bilingual teacher education, bilingual schools, and language research and training efforts through the INI have together effected a slowing in the numeric decline of Tarascan speakers. Work by scholars in the Instituto Michoacano de Cultura del Gobiemo del Estado and in the Colegio de Michoacán have resulted in growing interest, training, awareness, and standardization (of the alphabet, training materials, and of teaching methodology) for the survival of this linguistic isolate.

References

Argueta, Arturo. 2003. Instituto Nacional Indigenista, Perfil de Hablantes Indigenas, Estado de Michoacán. Web version at http://www.ini.gob.mx/monografias/purepechas.html

Boyd, Maurice. 1969. Tarascan myths and legends. Fort Worth: Texas Christian University Press.

Chamorro, Arturo. 1994. Emotion and rivalry in Purhepecha music. Paper presented at the 46th International Congress of Americanists, Stockholm, July 4–9.

Corona Nunez, Jose. 1957. La religi6n de los Tarascos. Morelia: Anales del Museo de Michoacán, No. 4.

Friedrich, Paul. 1975. A phonology of tarascan. Chicago: University of Chicago Press.

Gómez Bravo, Lucas, Benjamin Pérez González, and Ireneo Rojas Hernández. 2001. Intrudcción al idioma P'urhepecha. Morelia, Mich., Mexico: Universidad Michoacana de San Nicolas de Hidalgo, Centro de Investigaci6n de la Cultura P'urhepecha.

González, Juan L. Jr. 1985. Mexican and Mexican-American farm workers. New York: Praeger Publishers.

INI 2002 (source: http://207.248.180.194/bibdf/ini/perfiles/purhepecha).

Jasso Espinosa, Rodolfo. 1969. Grandeza de Michoacan. Mexico City: B. Costa-Amico.

Katzner, Kenneth. 1992. Languages of the World. New York: Routledge.

León, Nicolás. 1904. Los tarascos, notas hist6ricas, etnicas y antropo16gicas, Pt. 1. Mexico City: Imprenta del Museo Nacional.

Mendieta y Núñez, Lucio. 1940. Los Tarascos. Mexico City: Imprenta Universitaria.

Pollard, Helen. 1993. Tariacuri's legacy: the prehispanic Tarascan state. Norman: University of Oklahoma Press.

Relación de la ceremonias y ritos y población y gobierno de los indios de la provincia de Michoacán (1541.) 1956. Madrid.

Rubio Ortiz, Pascual. 1920. Historia de Michoacán. Morelia.

Ethnologue. 2003. Summer Institute of Linguistics (source: http://www.ethnoiogue.com/).

Velásquez Gallardo, Pablo. 1978. Diccionario de la lengua Phorhépecha. Mexico City: Fondo de Cultura Económica

West, Robert C. 1948. Cultural geography of the Tarascan area. Washington, DC: Smithsonian Institution, Institute of Social Anthropology.

WARREN D. ANDERSON

Teaching: Curricula

The term 'curriculum' refers to the content and aims of a course, the teaching procedures that will be followed, and the ways in which student learning will be assessed. For some, curriculum is synonymous with 'syllabus', that is, the listing and sequencing of course content. The terms 'syllabus' and 'method' are also sometimes used interchangeably. In its broadest sense, a curriculum includes all of these: syllabus, aims, method, and means of assessment.

A helpful way to summarize key issues in language curriculum design is through Richards and Rodgers' notions of approach, design, and procedure. 'Approach' refers to the theory of language and language learning that underlies the particular approach to course development. 'Design' includes the objectives, organization, and content of the particular syllabus type, kinds of teaching and learning activities, teacher and learner roles, and the role of instructional materials. 'Procedure' describes the classroom techniques and practices that might be employed within the particular approach.

For a long time, the overriding approach to language curriculum design was based on the structural syllabus; that is, the listing of grammatical items to be learned. This was the basis of the grammar-translation method, which dominated language teaching from the 1840s to the 1940s. Structural syllabi are still used today, but primarily with a different approach to methodology and in combination with other approaches to syllabus design.

Situational language teaching, or the oral approach, became the primary avenue toward language teaching in Britain in the 1950s. The aim of this approach was to develop learners' spoken language skills in everyday situations. Hence, lessons might be labeled 'eating out' or 'at the bank'. The view of language in this approach was still structural, however, but with a focus on oral practice and the use of sentence patterns and drills. The audiolingual method, which emerged in the United States at the same time, was similar in its view of language and use of methodology.

A major shift occurred, however, in the late 1960s, when applied linguists began to question the theoretical assumptions underlying situational language teaching and audiolingualism. Drawing on the work of scholars such as M.A.K. Halliday, Dell Hymes, John Langshaw Austin, and John Searle, what came to be known as the 'functional-notional' or the 'communicative' approach to language teaching emerged. This view saw the aim of language learning as involving more than just the acquisition of linguistic structures. The aim was the achievement of communicative proficiency and the ability to convey meaning in a range of communicative situations. This shift in the aim of language learning programs had a dramatic impact on language-learning curricula.

The basis of this new approach comprised the concepts of 'functions' and 'notions', based on the work of Austin and Searle, and of the applied linguist Wilkins, as well as the concept of communicative competence, based on the work of Hymes. Functions are the things that people 'do' with language, such as asking for information, greeting, apologizing, and inviting. Examples of notions are frequency, duration, location, and quantity. Communicative competence includes knowledge about language as well as methods of using language in appropriate ways to achieve particular goals. Language-learning programs were changed to make communicative competence, rather than (just) linguistic competence, the overall goal. As a result, lists of functions and notions and communicative appropriateness were added to the language-learning curriculum.

The communicative approach is still used today, but with a number of subsequent refinements and redevelopments. These include programs that focus on skills development (such as listening for 'gist', deciphering the main idea, and guessing vocabulary from context); programs that focus on tasks (such as exchanging information, solving problems, making decisions, reading a job advertisement, making an appointment, and writing a résumé); and programs organized around a particular content area (such as first year university science and technology) or themes (such as food, families, employment, and pollution). Other syllabus types include the 'lexical' approach, which focuses on the acquisition of particular vocabulary items, and text- or genre-based approaches, which use whole texts as the starting point for language program development and consider what it is that learners need to know and accomplish in order to produce an example of a particular genre. A further development is the 'negotiated syllabus', in which teachers and learners negotiate the

aims of the program, course content, and sometimes methodology. Often, however, language-learning curricula will be 'mixed' or 'multistranded', in that they will combine a number of these program types in order to be most useful and comprehensive for a particular group of learners.

James Dean Brown's *The elements of language curriculum* (1995*)* describes the main components of a language-learning program as needs analysis, goals and objectives, testing, materials, teaching, and evaluation. 'Needs analysis' refers to the identification of the language that students need in order to achieve particular goals, along with the identification of the learners' linguistic competence in relation to these goals. Needs analysis is an especially important component of programs that are designed for specific purpose language use, such as academic English and business English programs. 'Goals and objectives' refer to the overall reason or purpose of the program, as well as what learners are expected to be able to do with language by the end of the program. 'Evaluation' includes the ways in which learners' progress will be assessed in relation to the program's goals and objectives. Once needs have been established and goals and objectives have been set, the curriculum planner is able to go about adopting ready-made materials for the program, adapting materials as required, or developing new materials. Decisions then proceed regarding teaching approaches and program evaluation; that is, how the program might best be taught and how it can best be assessed in terms of its overall effectiveness.

References

Brown, James Dean. 1995. The elements of language curriculum, a systematic approach to program development. Boston and London: Heinle & Heinle.

Dudley-Evans, Tony, and Maggie Jo St. John. 1998. Developments in English for specific purposes. A multi-disciplinary approach. Cambridge and New York: Cambridge University Press.

Feez, Susan. 1998. Text-based syllabus design. Sydney: National Centre for English Language Teaching and Research, Macquarie University.

Hedge, Tricia. 2000. Teaching and learning in the language classroom. Oxford and New York: Oxford University Press.

Johnson, Robert (ed.) 1989. The second language curriculum. Cambridge and New York: Cambridge University Press.

Nunan, David. 1988a. Syllabus design. Oxford, New York: Oxford University Press.

———. 1988b. The learner-centred curriculum. Cambridge and New York: Cambridge University.

Richards, Jack, and Theodore Rogers. 1986. Approaches and methods in language teaching: a description and analysis. Cambridge and New York: Cambridge University Press.

Richards, J.C. 2001. Curriculum development in language teaching. Cambridge: Cambridge University Press.

Ur, Penny. 1996. A course in language teaching: practice and theory. Cambridge and New York: Cambridge University Press.

White, Ron. 1988. The ELT curriculum. Oxford and Cambridge, MA: Basil Blackwell.

Wilkins, David. 1976. Notional syllabuses. Oxford and New York: Oxford University Press.

Yalden, Janice. 1983. The communicative syllabus. Evolution, design and implementation. Oxford and Elmsford, NY: Pergamon.

BRIAN PALTRIDGE

See also **Genre; Second Language: Teaching; Teaching: Methods**

Teaching: Methods

A teaching method, sometimes referred to as a methodology, is a coherent and theory-based characterization of the language-teaching process, usually prescribing one set of classroom practices, procedures, strategies and techniques and often proscribing others.

A method is always informed by a broader approach or underlying philosophy. This would normally include a set of beliefs about language and how it works, about how people learn or acquire language, about curriculum design and general educational approaches, about general learning theories, and about the purposes of language learning.

Since there is general debate and disagreement about all of these broader issues, it is not surprising to find that the search for the 'best' language-teaching method has been fruitless. Studies that have sought to compare different methods to determine their effectiveness have remained inconclusive, partly because there are so many variables involved in the language teaching and learning process, partly because, as a number of empirical studies have discovered, the actual activities in observed classrooms often seem indistinguishable no matter what method is being used, and partly because it is almost impossible to ascribe cause and effect in anything other than trivial or short-lived learning.

The history of language-teaching methods has largely been one of fluctuation between form and function, between learning about the target language and learning through actually trying to use it, and between focus on the written language and focus on the spoken language. While there is little doubt that the vast majority of those people who successfully and fluently use more than one language—and there are many more bilinguals and multilinguals in the world than monolinguals—learned their languages outside the classroom by using them for real purposes in the real world, the issue of formal instruction is more complex. For example, language instruction in formal classroom settings has in many cases not included the purpose of learning to communicate in an everyday sense, but has instead focused on gaining access to the literary canon of the target language or has been for the specific purpose of understanding ancient texts for historical, cultural, or archeological reasons.

In the study of languages that are no longer spoken naturally, focus is necessarily on understanding the mechanics of the written language—the grammar and the meaning of words. The widespread study of Latin and Greek in western schools from the Middle Ages onward had a major influence on the teaching of languages more generally and meant that the grammar-translation method was the dominant—although not unchallenged—paradigm for language teaching until well into the twentieth century.

The major challenge came late in the nineteenth century when thinkers such as Henry Sweet in England and Wilhelm Viëtor in Germany argued strongly for the primacy of speech over writing, the inductive learning of grammar, and the avoidance of translation, using only the target language as far as possible. This ultimately led to the development of the direct method, also influenced by studies of natural language learning, especially in childhood. The direct method was readily adopted by private language schools, where native speakers of the target language were recruited as teachers. It was much less influential in formal education systems, where nonnative speakers often struggled with the fluency needed to implement the method, and where the emphasis on spoken language made formal assessment of students much more difficult.

The audiolingual method, the development of which owed much to the wartime need for fluent and convincing speakers of other languages, was a building-block approach based on a thorough analysis of language grounded in structuralist traditions and on the theories of behaviorist psychology, with its emphasis on stimulus–response conditioning and learning through habit formation. It became the dominant language-teaching method by the 1960s despite the fact that its theoretical foundations were already being severely eroded through the development of transformational grammar and Noam Chomsky's 1959 attack on habit formation as a basis for language learning. Many activities in the language-teaching classroom today still show allegiance to audiolingual and behaviorist principles.

The dominant paradigm over the last 30 years in language-teaching worldwide has been communicative language teaching. The approach foregrounds the use of language as a means of communication, with the assumption that much learning of language takes place through language in use rather than language as an object of study. Communicative language teaching was reinforced by advances in error analysis and second language acquisition studies in the 1970s, which strongly suggested that learners proceed developmentally through a predictable series of stages in which the production of errors plays an important part. It is also closely linked with the notion of language in context and with those sociolinguistic theories, especially M.A.K. Halliday's systemic-functional grammar and Dell Hymes' notion of communicative competence, that take context into account in their investigations of how people use language. The key concept in communicative language teaching is that learners learn through the use of language in meaningful activities. It is not simply a matter of contextualization of the language—it must be use of language that is meaningful for the learners.

The development of language for specific purposes, where language learning occurs while the ostensible focus is on communicating for some other purpose, can be seen as a special case of communicative language teaching. Similarly, Stephen Krashen and Tracy Tyrell's natural approach (1983), originating in studies of English-speaking Canadian students studying in French-medium schools ('immersion'), can be classified as communicative language teaching. Both of these have been very influential at a policy level, and there has been considerable work on language across the curriculum and on bilingual and immersion education.

An allegiance to the principles of communicative language teaching has shown up in national curricula in recent times, for example, in countries such as China, Korea, Japan, and Thailand; however, there are obstacles to implementation in practice, and communicative language teaching has been criticized for conflicting with local cultural values and for its unrealistic expectations of the teacher's own language proficiency.

The very concept of a language-teaching 'method' has also come under attack recently, with several theorists and researchers finding that the concept has little justification in actual classroom practices. Jack Richards (1990) has suggested that teachers would be better advised to focus 'not on the search for the best method, but rather on the circumstances and conditions

under which more effective teaching and learning are accomplished'.

References

Brown, H. Douglas. 1987. Principles of language learning and teaching. Englewood Cliffs, NJ: Prentice-Hall.

Candlin, Christopher N., and Neil Mercer (eds.) 2001. English language teaching in its social context. London and New York: Routledge.

Chomsky, Noam. 1959. A review of B.F. Skinner's verbal behavior. Language 35. 26–58.

Holliday, Adrian. 1994. Appropriate methodology and social context. Cambridge and New York: Cambridge University Press.

Howatt, A.P.R. 1984. A history of English language teaching. Oxford: Oxford University Press.

Krashen, Stephen D., and Tracy Terrell. 1983. The natural approach: language acquisition in the classroom. Oxford: Pergamon.

Richards, Jack C. 1990. The language teaching matrix. Cambridge: Cambridge University Press.

Richards, Jack C., and Theodore S. Rodgers. 1986. Approaches and methods in language teaching. Cambridge and New York: Cambridge University Press. 2nd edition 2001.

DAVID R. HALL

See also **Acquisition; Halliday, Michael Alexander Kirkwood; Hymes, Dell H.; Sweet, Henry; Teaching: Curricula**

Telugu

Telugu is a Dravidian language, one of the language families found in South Asia. The Dravidian family is divided into three subgroups: South Dravidian, Central Dravidian, and North Dravidian. Telugu belongs to the Central Dravidian subgroup. Some other languages of this subgroup—Gondi, Konda, Pengo, Manda, Kui, Kuvi—have sometimes been treated as a separate group called South-Central Dravidian. Kolami, Naiki, Parji, and Gadaba are the other languages in Central Dravidian.

Telugu is mainly spoken in the State of Andhra Pradesh in the south central part of India. Telugu is also spoken in the neighboring states—Tamil Nadu, Karnataka, Maharashtra, and Orissa. It is spoken by migrants in South Africa, Mauritius, Fiji Islands, Malaysia, Singapore, and the United States of America. The earliest written records of Telugu from the second century BC consist mainly of records of names. Numerous inscriptions written in Telugu are available from the sixth century AD. From the twelfth century onward, literary and grammatical treatises started being composed. On the basis of the available materials, some scholars have divided the history of the Telugu language into six periods:

(1)	Early Old Telugu	AD 200–AD 700
(2)	Late Old Telugu	AD 700–AD 1200
(3)	Early Middle Telugu	AD 1201–AD 1400
(4)	Late Middle Telugu	AD 1401–AD 1600
(5)	Early Modern Telugu	AD 1601–AD 1900
(6)	Modern Telugu	AD 1901–onward

Regional variations give us the following picture of dialects and subdialects:

(1) Coastal Dialect	—	Districts of Andhra Pradesh
(i) North Coastal	—	Vijayanagaram Vishakapatanam Srikakulam
(ii) Central Coastal	—	West Godavari East Godavari Krishna Guntur
(iii) South Coastal	—	Prakasam Nellore
(2) Rayalaseema Dialect		
(i) Southern Rayalaseema	—	Chittoor
(ii) Central Rayalaseema	—	Cuddapah Kurnool
(iii) Western Rayalaseema	—	Anantapur
(3) Telangana Dialect		
(i) Central Telangana	—	Medak Ranga Reddy Hyderabad
(ii) North Telangana	—	Adilabad Karimnagar Nizamabad
(iii) South Telangana	—	Mahaboobnagar
(iv) East Telangana	—	Warangal Nalgonda Khammam

Telugu shares the presence of cerebral consonants (ṭ, ṇ, Ḷ) with the other Dravidian languages. These consonants are not present in any of the Indo-European groups (the tongue position differs slightly from that used to produce English [d, n, l]) excepting in the Indo-Aryan subgroup, which perhaps is due to the influence of Dravidian languages.

Similarly, Telugu has long and short varieties of central vowels (e, ē, o, ō), which again is true in other Dravidian languages, but not in Indo-Aryan languages (e.g. erugu (to know), ērugu (to shit), toḍa (thigh), tōḍa (with)).

Telugu draws a distinction between voiced and voiceless consonants (e.g. /d/ vs. /t/), which is not the case in e.g. Tamil (a South Dravidian language). In Tamil, b, d, g, ḍ, j occur only in postnasal positions—i.e. after *n* or *m*—whereas p, t, k, ṭ, c occur word initially or in clusters. In Telugu, the following word pairs show that the difference between k/g or t/d is crucial: kampa (thorny wood) vs. gampa (basket), āṭa (play) vs. āḍa (feminine), kanta (hole) vs. kanda (an edible root).

Aspiration of consonants is also a contrastive feature that Telugu has acquired from Sanskrit and built into its orthographic system (e.g. d vs. dh); its orthographic system is derived from the Ashok Brahmi script. Speakers of nonstandard varieties of Telugu, however, tend to drop aspiration in their speech.

'Friction' (of voiceless labiodentals) is a feature that Telugu has recently borrowed from English but it is not integrated into its script, e.g. coffee (kāfī), sofa (sōfā). It is represented in the script by bilabial voiceless aspirate, i.e. the sound [f] is written as *ph*.

Sandhi and vowel harmony are reflected in the orthography that is phonemic (letters stand for individual speech sounds) and syllabic (letters stand for syllables). Keeping in mind the behavior of sounds in Sandhi, Telugu consonants can be classified as follows:

(1)	Dorsals	k, kh, g, gh
(2)	Apicals	ṭ, ṭh, ḍ, ḍh, n, ṇ, l, Ḷ, r
(3)	Laminals	t, th, d, dh, č, čh, ǰ, s, ś, ṣ
(4)	Labials	p, ph, b, bh, m

t/d and č/ǰ are regarded as belonging to one category (laminal), as examples of Sandhi (microjuncture) phenomena show: a phonological rule of Telugu states that a short vowel gets deleted between consonants of the same category. This juxtaposes two consonants and causes one of them to assimilate to the other. For example:

cādu   +   cukka   →   c ā c c u k k a
[substance   [dot]
used as dot
on one's forehead]

If we had treated d and c as belonging to two different classes, the deletion of the final /u/ as well as the subsequent assimilatory change from /d/ to /c/ (to match the following /c/) could not be treated as an instance of the broader Sandhi rule.

Vowel harmony in Telugu affects noninitial syllables:

enimidi   +   lu   →   enumudulu
'eight'   pl

All vowels in the first word assimilate (are harmonized) to the vowel in the suffix *-lu*. This kind of harmonization does not cross a geminate (doubled)_ consonant or a consonant cluster:

engili   +   lu   →   engiḶḶu
[partly   [pl]   (remnants of
eaten food]   eaten food)

Telugu, like other Indian languages, is a postpositional language, i.e. particles equivalent to English prepositions (such as *with*) follow the noun or noun phrase:

atanu   →   atani   tō
[that he   his   with (= with him)
nonpolite]

Nouns and pronouns that do not end in 'u' do not have distinct oblique form:

rangaḍu   →   rangaḍi tō
Ranga   Ranga with
āme   →   āme tō
[that she   her with (with her)
polite]
reḍḍi   →   reḍḍitō
Reddy   Reddy with (with
                                 Reddy)

In Telugu, the finiteness is expressed by the GNP (Gender Number Person) marker, which comes at the end of a verbal expression. If the GNP marker is absent, the verbal unit is nonfinite.

āme   ninna   vacci   īrōju   veḶḶindi
She   yesterday   come-perf.   today   go-perf.-she
She came yesterday and went (back) today

Here the marker '-i-' in *vacci* and *veḶḶindi* shows completion of action—hence 'perfective' suffix. 'ninna' and 'īrōju' refer to time. As the action is complete, people consider this also to be past tense.

As in the case of other languages, Telugu kinship terms reflect certain cultural phenomena like the

'marriage' relationship. In English, an 'uncle' could be maternal or paternal. Telugu has different terms for them. The maternal uncle, the father-in-law, and the paternal aunt's husband are all referred to as 'māma'. Similarly, the paternal aunt, the mother-in-law, and the maternal uncle's wife are referred to as 'atta'; both the atta's and the māma's daughter/son can become one's spouse.

In North India, where Indo-Aryan languages like Hindi and Punjabi are spoken, the maternal uncle, the father-in-law, and the paternal aunt's husband are given different terms because the maternal uncle's and the paternal aunt's son/ daughter cannot be taken as one's spouse.

Because Telugu words tend to end in a vowel, the language has been described by Charles Philip Brown, a great lexicographer, as 'the Italian of the East'.

References

Krishnamurti, Bh., and J.P.L. Gwynn. 1985. A grammar of Modern Telugu. Delhi: Oxford University Press.
Prakasam, V. 1985. The linguistic spectrum. Patiala: Punjabi University.
Ramakrishna Reddy, B. 1986. Localist studies in Telugu syntax. Hyderabad: The Author.
Ranganathacharyulu, K. 1987. A historical grammar of inscriptional Telugu (1401 A.D. to 1900 A.D.). Hyderabad: Dept of Linguistics, Osmania University.

VENNELAKANTI PRAKASAM

Tense and Aspect Marking

From the point of view of language typology, the most common grammatical means expressing tense and aspect are past, future, perfect (these three expressing tense) and progressive, imperfective, perfective (these three expressing aspect); cf. Dahl (1985), Bybee (1985), and Bybee and Dahl (1989). In the majority of cases, the present is morphologically unmarked, which is the reason why it is usually not regarded as a separate grammatical morpheme. The six tense–aspect expressions mentioned are found alone or in combination: e.g. in the case of the *pluperfect* (*John had seen the woman*, past plus perfect) or past progressive (*John was seeing the woman*, progressive plus past). Tense and aspect may be expressed either by periphrastic or by inflectional (or bound) expressions. Typologically, perfect and progressive have a very strong tendency for periphrastic expressions, while past, imperfective, and perfective prefer the bound variant. Future does not favor either periphrastic or inflectional expressions.

Tense is a grammatical category that specifies a time at which the corresponding tenseless sentence is true. Compare *I was often in Venice,* uttered at a certain time *t:* past expresses that there is a period of time before the speech time *t* at which the tenseless sentence *I often be in Venice* is true. Past indicates that the event at issue precedes speech time; future indicates that it follows speech time. Perfect (also called anterior) is used for events preceding speech time or some other reference time and still being relevant at that reference time.

Aspect is not relational like tense, but focuses the internal temporal structure of events. The progressive (also called continuous) indicates that an event is still in progress, relative to some explicit or implicit reference time, cf. *John was dancing (when Mary came in),* where the reference time is given through Mary's coming in. The perfective is used when the event is viewed as a bounded whole, whereas situations described in the imperfective are not seen as bounded. In Slavic languages, the imperfective/perfective distinction is mostly expressed morphologically, cf. Russian *pisal* 'to write' (imperfective) vs. *napisal* 'to finish writing' (perfective).

Now let us examine some common tense–aspect expressions in detail. As for the perfect, there are at least four different types of periphrastic constructions (cf. Dahl 1985): first, copula plus past participle (e.g. in Hindi, Bulgarian); second, auxiliary 'have' plus past participle (originally possessive constructions, e.g. in most Germanic and Romance languages); third, main verb plus particle 'already' (e.g. in Yoruba, Isekiri); fourth, constructions with auxiliaries historically developed from verbs meaning 'finish' or 'throw away' (e.g. in Sango, Ewe). The present perfect in English has the following uses (cf. McCawley 1971): (a) to indicate that a state of affairs prevailed throughout some interval stretching from the past into the present (universal perfect—*I've known Max since 1960*); (b) to indicate the existence of past events (existential perfect—*I have read Principia Mathematica five times*); (c) to indicate that the direct

effect of a past event still continues (stative/resultative perfect—*I can't come to your party tonight—I've caught the flu*); (d) to report hot news (hot news perfect–*Malcom X has just been assassinated*).

The *progressive* usually combines with the present and the past. Like the perfect, the progressive tends towards periphrasis. The periphrastic expressions are locative, meaning 'to be located in or at an activity'; cf. Bybee and Dahl (1989). Explicit variants (copula plus locative adposition plus nominalized verb) are most frequent, e.g. in Irish *tá sé ag dúnadh an dorais* ('he is at shutting the door'). Locative meaning is part of almost all progressives, even of the English progressive, although not overtly. But the English progressive like in *John is jumping* is historically derived from the construction *John is a-jumping*, where 'a' is actually the locative preposition 'on'. Although the progressive usually describes events that are still in progress at reference time, note that e.g. the English progressive has a wider use. In *John is writing his thesis*, the activity need not be in progress at reference time but rather at a larger time-interval properly including reference time. In addition, examples like *John is working on his thesis everyday* show that the progressive is also possible with habitual activities, at least in English.

The future does not favor either periphrastic or inflectional expressions. Common lexical sources for the future are the following: first, an auxiliary with the former meaning of 'want' (e.g. in Mandarin and Swahili, and also English 'will'); second, a verb meaning 'to owe' (e.g. in the Western Romance languages, Korean, and also English 'shall', derived from 'sceallan'); third, like the English going-to-future, a collocation meaning 'movement toward a goal'. The difference between the English will-future and the going-to-future is easily felt in minimal pairs like *there is going to be a riot here* vs. *there will be a riot here*. Only the going-to-future expresses that the riot is already in its preparation phase and on its way to happening. The will-future, in contrast, just makes a prediction, but the riot can also happen some weeks later, with no preparatory actions of it at the present moment.

The most widespread type of bound or inflectional tense-aspect is what Bybee and Dahl (1989) call the tripartite system. The past/nonpast distinction is added to the perfective/imperfective distinction, with perfective combining only with past, and the past/nonpast distinction only occurring in the imperfective. About every second language seems to have this tripartite system (e.g. Mandarin Chinese, French, Kurdish). Interestingly, the two distinctions are independent, because it is also possible to have the perfective/imperfective distinction but lacking past (e.g. in Arabic), or to have past but lacking the perfective/imperfective distinction, as in English. The tense–aspect systems of many Slavic languages are not simply tripartite, however. Russian, for example, allows present-imperfective, past-imperfective, past-perfective, future-perfective, and future-imperfective. The system of Ukrainian is even richer; in addition to the Russian possibilities, we also find perfect-imperfective, perfect-perfective, pluperfect-imperfective and pluperfect-perfective.

References

Bybee, J.L. 1985. Morphology: a study of the relation between meaning and form. Amsterdam: Benjamins.

Bybee, J.L., and Ö. Dahl. 1989. the creation of tense and aspect systems. Studies in Language 13.

Comrie, B. 1976. Aspect. Cambridge: Cambridge University Press.

———. 1985. Tense. Cambridge: Cambridge University Press.

Dahl, Ö. 1985. Tense and aspect systems. Oxford: Blackwell.

Hopper, P.J. (ed.) 1982. Tense—aspect: between semantics and pragmatics. Amsterdam: John Benjamins. .

McCawley, J.D. 1971. Tense and time reference in English. Studies in linguistic semantics, ed. by C.J. Fillmore and D.T. Langendoen. New York: Holt, Rinehart and Winston.

Tedeschi, P., and A. Zaenen (eds.) 1981. Tense and aspect. New York: Academic Press (= Syntax and Semantics 14).

MONIKA RATHERT

Tense: Syntax

Often, theories of tense try to assign a constant meaning to morphological tenses like future, preterite, or present. But a closer look at the data shows that one has to distinguish between morphological tense (small letters in this article) and semantic TENSE (capitals). Russian, for example, expresses a telic semantic FUTURE by a combination of morphological perfective plus morphological present. A German semantic FUTURE may be a morphological present.

While it is more or less known what the morphology of a morphological tense is, the semantics of a semantic TENSE is notoriously unclear. Some regard TENSES as operators (e.g. Ogihara 1996; Dowty 1979); others view them as constraints on variables (e.g. Partee 1973). The nature of the temporal relations is debated, too: most people regard TENSES as two-place operators, but there are also some (Nerbonne 1985; Ehrich 1992) modeling the three Reichenbachian parameters *E*, *R*, and *S* as a three-place relation. For most researchers, TENSES locate the time variable of the verb; others view them as restrictions on the possible denotations of this variable (Dowty 1982; Nerbonne 1985).

An interesting issue of morphological tense vs. semantic TENSE is given by the sequence-of-tense (SOT) phenomena. Now it is widely accepted that tense in attitude-complements may be ignored in the semantics, i.e. that tense in attitude-complements may be no TENSE, that there is TENSE-deletion or empty TENSE.

The interpretation of temporal expressions in embedded contexts is a complex issue; there are contrasts between temporal expressions in the scope of attitude verbs and those in adjunct clauses, and there is significant cross-linguistic variation (Kusumoto 1999; Ogihara 1996).

The semantic treatment of embedded tenses began with Enç (1986, 1987). According to her, tenses are to be considered referential expressions denoting times; they fill temporal argument positions. Tenses are subject to anchoring principles that are sensitive to GB-style binding restrictions. This works fine for matrix clause but, as pointed out by Ogihara (1996), it runs into trouble when applied to multiple embeddings. In these cases, a tense has to meet contradictory anchoring requirements:

(1) John decided a week ago that in ten days at breakfast he would say to his mother that they were having their last meal together

In (1), the time of their last meal is understood as simultaneous to John's speech (thought there is a past tense) and, considering the presence of the adverbial 'in ten days', this time is in the future with respect to speech time.

A semantically more sophisticated analysis is proposed by Ogihara (1996). He developed an analysis in which TENSES are Priorean operators, predicates take a time variable as their argument, and the simultaneous interpretation is accounted for by the application of a TENSE deletion rule. When a PAST TENSE is c-commanded by (embedded in) another PAST TENSE, the former is deleted and the embedded sentence becomes TENSEless:

(2) John said that Mary was pregnant
John PAST say that Mary PAST be pregnant
John PAST say that Mary $\varnothing$ be pregnant

Ogihara's system was also developed in order to account for the crosslinguistic contrast between English and Japanese. In Japanese, we find the semantically right TENSE configuration since simultaneity under a matrix PAST is expressed by the TENSEless 'present' and there is no need for the deletion rule:

(3) Bernhard-wa Junko-ga byookida to it-ta
B.-TOP Junko sick-$\varnothing$ that say-
-SUBJECT PAST
'Bernhard said that Junko was sick'

Moreover, Ogihara can account for the English–Japanese contrast of TENSE interpretation in relative clauses. A PAST TENSE relative clause embedded under a matrix PAST TENSE can have a simultaneous interpretation by the application of the deletion rule:

(4) John saw a girl who was running
John PAST see a girl who PAST be running
John PAST see a girl who $\varnothing$ be running

In Japanese, the simultaneity under a matrix PAST is expressed by the TENSEless 'present' and there is, again, no need for the deletion rule:

(5) Mariko-wa naiteiru otoko- hanasi
noko-ni kaketa
Mariko-TOP cry-teiru $\varnothing$ boy-to talk-PAST
'Mariko talked to the boy who was crying'

Expressions other than 'real' tenses but associated with PAST-TENSE-like interpretations also trigger TENSE deletion in Ogihara's system. TENSE deletion can be triggered by the perfect and by certain noun complements:

(6) John believes Mary to have claimed that she was innocent

(7) John's claim that he was innocent is well known

Ogihara accounts for the phenomena by assuming that a noun like *claim* can be associated with the syntactic features [+PAST], [–PAST], the perfect with the feature [+PAST], the past tense with [+PAST], and the present tense with [+PRES]. Under these assumptions, the SOT rule is simply sensitive to the syntactic features [+PAST] and [+PRES]:

(8) SOT rule: If a tense feature B is the local tense feature of a tense feature A at LF, and A and B are occurrences of the same feature, A and the tense associated with A are optionally deleted.

Ogihara's analysis is problematic for the following example, where the embedded future tense can only have a simultaneous interpretation:

(9) Jen   said that   would   move to
                 Cecilia   Amherst
     Jen PAST-say that   woll-PAST move to
                 Cecilia   Amherst
     Jen PAST-say that   woll-$\varnothing$ move to
                 Cecilia   Amherst

According to Ogihara, *would* has to be analyzed as composed of a TENSEless future operator *woll* and the PAST TENSE. Since the PAST TENSE on *woll* is c-commanded by the matrix PAST TENSE, it can be deleted, giving rise to the correct future-in-past interpretation. But since the TENSE deletion rule is optional, it should be possible to retain the embedded PAST, giving rise to a shifted future-in-past reading: there is a time *t* preceding Jen's speaking, and after this time (potentially before Jen's speaking), Cecilia moves to Amherst. This is clearly not a possible reading. In order to avoid the problem, Ogihara has to say that the SOT rule is obligatory in the case of *would*.

Since it is a syntactic configuration that triggers the application of the deletion rule, Ogihara's system cannot account for non-SOT languages such as Russian or Polish. In these languages, the present tense has different interpretations in complement and relative clauses when embedded under a matrix PAST.

References

Dowty, D. 1979. Word meaning and Montague grammar: Synthese Language Library. Dordrecht: Reidel.
———. 1982. Tense, time adverbs and compositional semantic theory. Linguistics and Philosophy 5. 23–55.
Ehrich, V. 1992. Hier und Jetzt - Studien zur lokalen und temporalen Deixis im Deutschen. Tübingen: Niemeyer.
Enç, M. 1986. Toward a referential analysis of temporal expressions. Linguistics and Philosophy 9(1). 405–26.
———. 1987. Anchoring conditions for tense. Linguistic Inquiry 18. 633–57.
Kusumoto, K. 1999. Tense in embedded contexts. Ph.D. Dissertation. Department of Linguistics, University of Massachusetts.
Nerbonne, J.A. 1985. German temporal semantics: three-dimensional tense logic and a GPSG fragment. Ph.D. Dissertation. The Ohio State University.
Ogihara, T. 1996. Tense, attitudes, and scope. Dordrecht: Kluwer.
Partee, B. 1973. Some structural analogies between tenses and pronouns in English. Journal of Philosophy 70. 601–9.

MONIKA RATHERT

Tewa and the Kiowa-Tanoan Languages

The Kiowa-Tanoan language family consists of four branches: Kiowa, Towa, Tiwa, and Tewa. Although they are not so radically divergent that they suggest a long time-depth to account for internal diversity, the speech communities associated with them and their component languages reside in two distinct cultural areas of Native North America. Kiowa has long been associated with the Southern Plains cultural area. This area is often associated with its cultural emphases on the tipi, the horse, the bison, and the sun dance ritual complex. In contrast, Towa, Tiwa, and Tewa—usually grouped together as the Tanoan languages—represent the Pueblo cultural area of the southwest. These cultures were known for their large, sedentary agricultural villages featuring elaborate ceremonial calendars. Anthropologists have located the Kiowas near the head of the Missouri River around 1700. While their earlier history is more difficult to construct, it is likely given both archeological and linguistic evidence that the Kiowas abandoned a more Pueblo-like adaptation in what is today Northern New Mexico well prior to Spanish exploration in the mid-sixteenth century, and began a process of cultural transformation that made them exemplars of the Southern Plains culture area. Part of the linguistic evidence for such an interpretation rests on the well-established fact that speech communities associated with every other branch of the Kiowa-Tanoan language family was associated with sedentary settlements in the Rio Grande River Valley of northern New Mexico at the time of Spanish contact in 1540.

Within the Tanoan group, the Towa branch consists of only one existing Pueblo community, Jemez Pueblo, although the extinct Pueblo of Pecos also spoke this language prior to abandonment.

The Tiwa branch is divided into Northern and Southern Tiwa languages, each having multiple dialects. In Northern Tiwa are the dialects of Taos and Picuris while Southern Tiwa consists of Sandia, Isleta, and Ysleta del Sur. The last of these dialects was formerly spoken in the Rio Grande area but transported by Isletans who migrated with the Spanish to El Paso in 1682 in the wake of the Pueblo Revolt of 1680. The Tewa branch consists of two languages—Rio Grande and Arizona Tewa. Rio Grande Tewa is or was spoken

in the following Pueblos—San Juan, San Ildefonso, Nambe, Santa Clara, Pojoaque, and Tesuque. In the last two of these Pueblos, the Tewa language has not continued except through the in-migration of speakers from other Tewa-speaking pueblos. Of the dialects currently maintained, Santa Clara is usually regarded as the most divergent. Arizona Tewa, which is currently spoken on First Mesa of the Hopi Reservation in northeastern Arizona, was formerly spoken by the Southern Tewa, also known as Tano, who occupied the Galisteo Basin of Northern New Mexico prior to their exodus from New Mexico after the (Second) Pueblo Revolt of 1696. Also known as the Hopi-Tewa or the Hano Tewa, the Arizona Tewa have continued to maintain their language for more than 300 years despite their relocation and accommodation to their more numerous Hopi neighbors.

Some Structural Features of Kiowa-Tanoan Languages

Most Kiowa-Tanoan languages have a large inventory of consonants, which includes a four-way contrast between voiced, voiceless, aspirated, and glottalized stops (e.g. b, p, ph, p'—where the voiceless bilabial is accompanied by a closure of the glottis as in the American English expression 'oh-oh'). The vowel system usually contains six vowel qualities but also adds vowel length and an oral/nasal contrast to expand the inventory. Some of the Kiowa-Tanoan languages have well-attested tonal contrasts. In Arizona Tewa, which has two tones, nouns pronounced identically in terms of consonants and vowels but contrasting in pitch may be different words—'shirt' (low tone) and 'digging stick' (high tone). The languages all have an elaborate verbal morphology. Verbs obligatorily consist of a pronominal prefix that codes the subject (and sometimes the object as well), the verb stem, and one or more tense–aspect suffixes. But verbs can also include prefixes that encode negation and adverbial manner information, incorporated objects (such as things owned or possessed), and suffixes indicating that the verb is part of a dependent clause rather than the main sentence. The intricate pronominal prefixes of these language encode the person (I, you, he/she/it) and number (singular, dual, and plural (>2)) of the subject and inflect for stative and possessive constructions. Other prefix sets supply pronouns for transitive sentences which often distinguish between agentive, reflexive-reciprocal, and inverse constructions. The last of these provides a means of making the recipients of action rather than its agents the subject of the sentence. As such, they resemble passive sentences in English, but unlike English passives, Tewa inverse constructions remain fully transitive. The most

common basic word order in the language appears to be Subject–Object–Verb and all languages accordingly feature postpostional phrases (e.g. house-into) rather than prepositional phrases.

Historical Linguistics—An Overview

The relatively close relationship of the Tanoan languages was first recognized by John Wesley Powell in 1891 on the basis of a lexical comparison of linguistic data that he had collected. J.P. Harrington was the first to link this group to Kiowa in 1910 when he published a list of likely cognate words they shared. Further comparative evidence was amassed until the relationship was definitively established by Kenneth Hale in 1967 when he successfully reconstructed the phonology of Proto-Kiowa-Tanoan—the 'protolanguage' or antecedent ancestral language, which precedes the emergence of the Kiowa, Towa, Tiwa, and Tewa branches.

Although Kiowa is the most recently detected member of the family, it is not necessarily the most divergent. The apparent divergence is greatly magnified by Kiowa's relative removal from recent contact with other related languages and by linguistic adaptation to a rather different cultural area. Although the Tanoan languages clearly share many more vocabulary items, formal resemblances in grammar crosscut the family and make Kiowa appear more like just another related language rather than like one that is marginally related.

Researchers have posited a deeper historical linguistic relationship that would relate the Kiowa-Tanoan language family to the Uto-Aztecan family of North and Meso-America. The linguist Edward Sapir, in his classifications of 1921 and 1929, was the first to posit such a relationship and this largely impressionistic assessment of verbal morphology later led to a more systematic examination of possible cognate lexical items by George Trager and Benjamin Whorf in 1937. Since that time, collective historical linguistic scholarship has undermined anything that might approximate a 'proof' of relationship and suggested that existing research is at best inconclusive. At the present time, this posited relationship must await new research before it can be definitively confirmed or disconfirmed.

Contemporary Speech Communities

Like virtually all indigenous languages that do not enjoy state support, many of the existing Kiowa-Tanoan languages are spoken by relatively small populations in which speakers are not evenly distributed across all generations. For the Kiowa, who now live primarily in southwestern Oklahoma in Caddo, Kiowa, and Comanche counties, there are about 300 adult

speakers exhibiting a wide range of fluency. Towa, spoken at Jemez, is reported as spoken throughout the community of about 2,000. The Northern Tiwa-speaking community of Taos, according to surveys conducted in 1980, has about 800 speakers across all generations and represents, after Jemez, the next most vital of Kiowa-Tanoan speech communities. Picuris has slightly more than 100 speakers. Of the Southern Tiwa dialects, Sandia has about 150 while Isleta has ten times that number. The Ysleta del Sur dialect is not fluently spoken anymore. Some members of the community remember some of the distinctive vocabulary of their ancestral language. As for Tewa, estimates based on surveys of various kinds count 1,200 speakers of Rio Grande Tewa and about 300 for Arizona Tewa. In both communities, there are children still learning the language, although the percentage still learning the language in their homes may be slightly less than 50%. Language renewal programs of various types exist in Southern Tiwa, San Juan Tewa, and some other communities as resources used by these native communities in their ongoing efforts to control and maintain their heritage languages.

Language and Culture

In all Kiowa-Tanoan speech communities, as in Native American communities more generally, speech is highly regarded as a powerful form of action and not reduced, as in some Western folk views, to labeling functions that linguists call 'denotation'. In all the Kiowa-Tanoan-speaking communities, traditional storytelling was viewed as an essential aspect of language socialization. In these stories children would experience the early moral training of learning from the positive and negative models provided by the story characters. In all communities, lively and engaging performances—typically by parents and grandparents—called attention to the dramatic power of language to entertain and instruct. All narrators use 'hearsay' grammatical devices to traditionalize and create a proper sense of authenticity for their narratives. Storytellers in the Kiowa tradition tell their stories collaboratively with their audiences. They expect considerable audience involvement as both narrators and audience members openly comment on the story as it progresses. In the Pueblo cultures of the Tanoan communities, audience involvement is more limited to stylized interjections that encourage the storyteller to continue that the narrative and that demonstrate the involvement of the audience. Kiowa narrators especially value humor and use many puns toward that end, whereas the verbal artistry of Tanoan storytellers is more centered on unfolding their stories through the dialogue of their characters.

While linguistic and speech-making skills were an important resource associated with tribal leadership among the Kiowa, the role of cultural models of speaking provided by the theocratic elite is especially pronounced in Tanoan communities. In many of these communities, the religious language of the kiva—the ceremonial chambers of native religious activity—provided a powerful, ideal model for everyday speech. Since this ceremonial register embodied both religious authority and the political power of the religious elite, it provided a natural symbol of good and effective speech, which accounts for the usual characterization of Pueblo cultures as 'linguistically conservative'. Such a cultural standard provides a model analogous to the standardized languages of state societies, which enjoy a high prestige, in part, because they are the official languages of state-owned and controlled institutions like public education and the broadcast media.

Some of the ideals embodied by kiva speech, which have had a demonstrable effect on speakers of Tanoan languages, are its emphasis on indigenous purism and the valorization of traditional forms. In the ceremonial speech of the kiva, no mixing of language is permitted; such mixing is in fact considered a punishable offense. Given this cultural ideal of linguistic purity, it is not surprising to find well-attested patterns in all the Tanoan languages of rejecting possible loanwords in favor of extending or combining existing terms. For example, the first word for automobiles among the Arizona Tewa was not a loanword from English 'car' or 'auto' but rather the word wa-tege (wind-wagon). For both Rio Grande and Arizona Tewa, patterns of linguistic purism have been well documented by Edward P. Dozier (Santa Clara Tewa) and Paul V. Kroskrity. Although bilingualism in Spanish for the Rio Grande Pueblos is traceable to the colonial occupation of the seventeenth century, less than 5% of the Rio Grande Tewa vocabulary is from Spanish, and for the Arizona Tewa who actively removed themselves from Spanish domination around 1700, less than 20 Spanish loanwords are attested. For the Arizona Tewa, this pattern of resistance to loanwords has also extended to their Hopi neighbors. Despite bilingualism in Hopi, pervasive intermarriage, and other forms of cultural integration, they only have two Hopi loanwords. In a similar fashion, the ceremonial emphasis on reproducing traditional forms as precisely as possible extends to a cultural value on traditionalizing narrative and other speech genres so that speakers lend their voices to the tradition of 'speaking the past'.

References

Davis, Irvine. 1979. The Kiowa-Tanoan, Keresan, and Zuni languages. The languages of native America: historical and comparative assessment, ed. by Lyle Campbell and Marianne Mithun, 390–443. Austin: University of Texas Press.

Kroskrity, Paul V. 1993. Language, history, and identity: ethno-linguistic studies of the Arizona Tewa. Tucson: University of Arizona Press.

Kroskrity, Paul V. 1998. Arizona Tewa Kiva speech as a manifestation of a dominant language ideology. Language ideologies, practice and theory, ed. by Bambi B. Schieffelin, Kathryn A. Woolard, and Paul V. Kroskrity, 103–23. New York: Oxford University Press.

Palmer, Gus. 2003. Telling stories the Kiowa way. Tucson: University of Arizona Press.

Trager, George L. 1946. An outline of Taos grammar. Linguistic structures of Native North America, ed. by Harry Hoijer et. al., 184–221. Viking Fund Publications, No. 6. New York: Viking Fund.

Trager, George L. 1967. The Tanoan settlement of the Rio Grande area: a possible chronology, ed. by Dell H. Hymes and William E. Bittle, 335–49. The Hague: Mouton.

Watkins, Laurel J. 1984. A grammar of Kiowa. Lincoln, Nebraska: University of Nebraska Press.

<div align="right">PAUL V. KROSKRITY</div>

Text Linguistics

In the 1960s, widespread dissatisfaction with prevailing models of linguistics, particularly in continental Europe and the United Kingdom, led to the rise of 'text linguistics' as a new linguistic subdiscipline. At that time, phonetics/phonology, morphology, and syntax were considered to be the core fields of linguistic inquiry. Furthermore, most linguists restricted their attention to the structure of the language system (*langue*) at the expense of language in use (*parole*). '*Langue*' is the idealized knowledge of a language that a competent speaker has. '*Parole*' refers to the less than perfect utterances this knowledge results in, characterized, for instance, by errors, repetitions, and hesitations due to processing constraints or not paying attention. The language system was thought to provide a window into the human mind, while 'language in use' was deemed too unsystematic and limited to be of real theoretical interest.

Consequently, the two central concerns of early text linguists were (a) linguistic modeling above the level of the sentence, and (b) a focus on language in use. The relationship between these two concerns has changed over time, however. Initially, most text linguists wanted to show only that structural properties extended beyond the level of the sentence. In the 1960s, Peter Hartmann, a German linguist, felt that the emphasis on phonemes, morphemes, and sentences was misplaced, as none of these elements alone constituted communication. Communication only occurs when these units are joined together in a text and, consequently, he viewed the text as the primary linguistic sign ('*das originäre sprachliche Zeichen*'). The search for a 'text grammar' or 'text syntax', therefore, abounds in this early work. However, it soon became apparent that textual rules do not have the same status as syntactic rules. The flouting of a syntactic rule results in an ungrammatical sentence. For instance, a syntactic rule of English is the categorical subject–verb–object order as in *The baby* (S) *loves* (V) *her new toy* (O). Violating that order by switching the constituents around as in *Loves* (V) *her new toy* (O) *the baby* (S) produces an ungrammatical sentence. However, textual rules may easily be flouted in certain contexts or for stylistic reasons and the text will still remain acceptable. For instance, a textual rule would be that pronouns normally appear after the element they refer to, as in *The baby loves her new toy. It makes such a funny sound*, where 'it' refers back to 'her new toy'. However, it is easy to image a context in which it would be perfectly acceptable to change the order around: *It makes such a funny sound. The baby loves her new toy.*

At that point, text linguistics began to examine the *langue–parole* distinction more critically. This was not a completely novel undertaking, as functional linguists had always been wary of the distinction. Functional linguists of the Prague School (founded as the '*Cercle Linguistique de Prague*' in 1926 by Villem Mathesius, Joseph Vachek, and others), for instance, had always maintained that language can only be described on the basis of authentic speech samples. As a result, they had also dismissed the strict distinction between *langue* and *parole*. In addition to the Prague School of functional linguistics, an important British group of functional linguists centered around John Firth, Michael Halliday, and John Sinclair, a recent offshoot of which is the Sydney School. As a consequence of the developing interest in texts, functional linguistics—a marginalized type of linguistic inquiry—gained in importance. It became increasingly apparent to text linguists that the only way to investigate textual structures was by turning to corpora of naturally occurring language. Text linguists also became increasingly interested in the patterns of

speech (initially, 'texts' had been thought of mainly as written language).

In 1981, Robert de Beaugrande and Wolfgang Dressler developed a set of criteria in an effort to define the characteristics of text as both (a) a unit of language above the sentence level and (b) language in use. These seven criteria are cohesion, coherence, intentionality, acceptability, informativity, situationality, and intertextuality.

First and foremost, texts are characterized by internal links. The word 'text' derives from Latin 'textus', which means '(inter)woven'. Grammatical links are known as 'cohesion' and semantic links as 'coherence'. Cohesion may be achieved through repetition, pronouns, ellipsis, conjunctions, and similar devices. However, textual linkage does not suffice to make a text. In the following example, all the links are present: 'I went to the library. Libraries contain books. Big volumes are heavy'. Despite the fact that 'library' and 'libraries' are linked through repetition, and 'books' and 'volumes' through synonymy, it is difficult to image a context where this would be an appropriate text. The textual characteristics of 'intentionality' and 'acceptability' refer to the fact that it is essential that a sender (the speaker or writer) intends to produce a well-formed text, and that addressees (hearers or readers) are prepared to accept the text as a communication that makes sense to them. Both senders and addressees usually rely on their encyclopedic knowledge (knowledge they have about the world that is not encoded in a particular text), in order to encode and decode texts. 'Informativity' refers to the relationship between old and new information that a text contains. A text containing only old information is dull; a text containing only new information becomes impossible to process. Therefore, senders produce texts that are in line with what they believe their addressees will already know, or not yet know. 'Situationality' refers to the factors that make a text appropriate to a particular situation. Finally, texts are 'intertextual' because they depend upon a knowledge of other texts to be understood. Such intertextual knowledge may refer to text types (we have certain expectations about what a letter looks like, or an encyclopedia entry) or to allusions (the slogan 'Just say know' as used in a students' union campaign alludes to the antidrug campaign 'Just say no', and the intertextuality forms an additional layer of meaning).

As the two understandings of text (as linguistic unit above the level of the sentence, and as language in use) became increasingly integrated, the need for a grammatical model able to incorporate the context has risen in importance. Halliday and other systemic functional linguists suggest that textual variation forms patterns along three dimensions of the context: field, tenor, and mode. 'Field' refers to context as social action. For instance, if a text about text linguistics is written for an encyclopedia entry, the language used will differ significantly from a text about the same topic that is produced as so-called 'cocktail party talk'. 'Tenor' refers to the role structure that pertains between the interactants, i.e. a text that is produced as a student–teacher exchange will differ from one that is produced as a marital communication. Finally, 'mode' refers to the symbolic organization of the text (e.g. written vs. spoken language).

The increasing focus on 'language in use' has blurred the boundaries between text linguistics and other fields of linguistics that share this concern, such as conversation analysis, discourse analysis, the ethnography of communication, interactional sociolinguistics, or pragmatics. Even if it seems likely that discourse analysis will eventually fully absorb text linguistics, text linguistics has to be credited with opening up new horizons for linguistics—by looking beyond the sentence, by trying to theorize context as part of grammar, and by opening the doors for interdisciplinary cooperation with education, literary studies, and other fields.

References

Brinker, Klaus, Gerd Antos, Wolfgang Heinemann, and Sven F. Sager (eds.) 2000. Linguistics of text and conversation: an international handbook of contemporary research, Vol. 1. Berlin and New York: Mouton de Gruyter.

de Beaugrande, Robert. 1997. New foundations for a science of text and discourse: cognition, communication, and freedom of access to knowledge and society. Norwood, CT: Ablex.

de Beaugrande, Robert. 1980. Text, discourse and process. London: Longman.

de Beaugrande, Robert, and Wolfgang Dressler. 1981. Introduction to text linguistics. London: Longman.

Eggins, Suzanne. 1994. An introduction to systemic functional linguistics. London: Pinter.

Halliday, M.A.K., and Ruqaiya Hasan. 1976. Cohesion in English. London: Longman.

Harweg, Roland. 1968. Pronomina und Textkonstitution. Heidelberg: Fink.

Martin, James R. 1992. English text: system and structure. Amsterdam and Philadelphia, PA: Benjamins.

Petöfi, Janos S. (ed.) 1988. Text and discourse constitution: empirical aspects, theoretical approaches. Berlin and New York: Mouton de Gruyter.

van Dijk, Teun A. (ed.) 1977. Text and context. London: Longman.

———. 1978. Textwissenschaft. Tübingen: Niemeyer.

Wilson, John, and Srikant Sarangi (eds.) 2000. Text linguistics at the millennium: corpus data and missing links. Special edition of Text, 20:2. Berlin and New York: Mouton de Gruyter.

INGRID PILLER

See also **Coherence in Discourse; Discourse Analysis; Firth, John Rupert; Genre; Halliday, Michael Alexander Kirkwood; Relevance in Discourse; Stylistics**

Text Understanding

Text understanding (Grosz et al. 1989; Scha et al. 1987; Singer 1990) is a cognitive process in which a text, i.e. a written sequence of 'coherent' sentences, is read and the human reader is able to decode all or most of the propositions encoded in that source. (In case the reader is a computer, we refer to this process as automatic text understanding (Schank and Christopher 1981)). Determining these propositions is, first of all, a challenging *semantic/conceptual* recognition problem. Because texts such as newspaper articles, cooking recipes, or business letters contain a large number of propositions, two additional layers help to structure the textual information flow. One is constituted by *structural organization* patterns of texts themselves, which reflect several mechanisms concerning how sets of text propositions are properly 'packaged'. Conditions of textual well-formedness can then be stated in a 'text' grammar much in the same way as a standard sentence grammar reflects sentential and phrasal well-formedness criteria. The second important issue concerns the *pragmatic* factors underlying text understanding, i.e. assumptions both writers and readers rely on (such that certain beliefs are shared or implicit goals can be recognized) in order to discern the *intended* (not only the literal) interpretation of a text. Therefore, we will consider text understanding from these three different angles, viz. the semantic/conceptual, structural, and pragmatic level (Grosz and Candace 1986).

In order to decode the contents encoded in a text, linguistic knowledge from various sources must be combined—knowledge of the morphological structure of the words and their organization in the lexicon, as well as grammatical knowledge in terms of syntax and semantics. Rather than just extending truth-functional semantic sentence analysis up to the text level (Kamp and Uwe 1993), text understanding—and hence the computation of text propositions—is deeply rooted in the knowledge of the textual domain. This background knowledge has to be made explicit in terms of a domain knowledge base, a repository that contains common-sense knowledge, as well as fairly specialized domain knowledge, e.g. from science, sports, or economy.

As far as the organization of that knowledge is concerned, research in text comprehension indicates that it seems beneficial to have structured knowledge representation formats (such as frames, scripts, or semantic networks) available. Such representation structures offer a densely packed interpretation context with taxonomic concept hierarchies, as well as default properties and predictive expectations that guide the recognition and incremental interpretation of objects and events. Text understanding processes operate on facts and basic propositions comprised of this a priori knowledge, either by finding a direct match with given knowledge structures, or by inferring a proper linkage, i.e. by employing reasoning mechanisms in order to link single propositions from a text to that background knowledge (Schank and Christopher 1981). Most of these inferences are merely plausible (e.g. inductive, abductive, uncertain, or default-based) rather than strictly deductive. As a result of incrementally assimilating knowledge from texts, one acquires entirely new propositions, modifies already known propositions, or even retracts propositions completely.

Text propositions are considered to be basic conceptual units such as predicate-argument pairs. For instance, in 'Peter met John on the beach. They spent three hours together.', we distinguish a MEETING-1 event whose actors are PETER and JOHN, whose location is ON-BEACH and whose duration is 3-HOURS. From this, we may construct the following predicate-argument representations as the main propositions:

MEET-ACTORS (MEETING-1, {PETER, JOHN})
MEET-LOCATION (MEETING-1, ON-BEACH)
MEET-DURATION (MEETING-1, 3-HOURS)

Even this small text fragment already features a pattern of text structure that guides our interpretation of both sentences. We may only derive the third proposition (specifying the duration of the meeting) when we interpret the pronoun '*they*' at the beginning of the second sentence as denoting '*Peter and John*' as they meet on the beach. This is an example of a reference due to anaphora: an anaphoric expression ('*they*') referring to some referents previously introduced in the text (the antecedent)—in this example, the ad hoc group of people is composed of '*Peter and John*'. Anaphoric reference is one of the basic *text cohesion* mechanisms that create local connections between adjacent sentences at the semantic/conceptual level.

In order to account for such cohesion phenomena, special representational devices for the history of a discourse (such as focus stacks, focus spaces, or centering lists (Walker et al. 1988)) must be available. They help to keep track of the attentional structure of

a text, e.g. by recording the kinds of specific discourse entities mentioned in the text, the order and syntactic embeddings in which they were introduced, etc. Such a device makes explicit the context created by the various utterances in a text in terms of the salience status of discourse entities and, thus, helps to minimize the number of alternative interpretations (ambiguities) that are likely to otherwise arise.

Text cohesion mechanisms are supplemented by *text coherence* patterns that relate to the global organizational structure of a text. Examples of this are various forms of coherence relations, e.g. causality, contrast, or elaboration (Mann and Sandra, 1988), hierarchically linking simple, as well as more and more complex proposition sets from the text. Moreover, text macroorganization patterns such as genre-specific superstructures or story grammars constrain the overall structure of the entire text (e.g. the composition rules underlying a weather report or a hospital discharge summary) (van Dijk 1977). Discourse structures of this type can be signaled by cue phrases (such as 'but', 'for example', 'however') and are, therefore, easily recognized, but they may also require elaborate inferencing on rich knowledge structures.

At the *pragmatic* level (Cohen et al. 1990), text understanding relies on discourse constraints that hold irrespective of particular utterances in a text and characterize a generally valid framework of human communication. They require the representation and recognition of various presuppositions (knowledge assumed to be mutually shared between the writer and the reader), goals, plans, and intentions of the writer, and adherence to truly general communication principles based on plausibility and rationality (such as several conversation maxims). These underpinnings of text understanding are intended to keep it focused by relying on a set of reasonable, socially accepted conventions for felicitous and compact textual information transfer. The pragmatic level of text understanding implies that utterances in a text contain more than they literally indicate in terms of semantic/conceptual information (a phenomenon called *conversational implicature*). The reader's construal of the implicit intentions, plans, and beliefs that the writer had for producing the text constitutes the key for a successful textual information transfer (Morgan and Georgia 1980).

A problem with defining text understanding in terms of the recognition of text propositions is that it is almost impossible to reach a consensus among different readers as to what exactly constitutes the proper set of propositions of an underlying text. Because this makes it difficult to judge whether a reader has or has not understood a text, several proposals have been made concerning how to assess the depth and accuracy of text understanding in a more or less indirect way. The following tasks have been considered useful in reaching that goal:

- In the *paraphrasing* task, subjects are asked to rewrite the contents of a source text in their own words, while keeping in mind any piece of information they have acquired from that source text. By way of comparing the propositions of the source text with those from the newly generated paraphrase text, the level of detail of text understanding can be made explicit.
- In the *question answering* task, subjects are asked particular questions whose answer is contained only in the source text. The answer to a question may either be explicitly stated in the underlying text, or it may have to be inferentially derived when it is merely implicit in the proposition set. Correctly answering a question is then taken as an indication of profound understanding of the source text.
- In the *summarizing* task, subjects are asked to compress the contents of a source text with respect to the most relevant propositions contained in it. Psycholinguists have a clear understanding of how one moves from a large set of basic propositions (*micro*propositions) to a generalized set of so-called *macro*propositions, which is the equivalent of a summary at the representation level. Transformations operating on micropropositions generate macropropositions by way of generalizing single statements along taxonomic concept hierarchies, in turn eliminating irrelevant details and abstracting single actions in more general event patterns (Kintsch and Teun 1978). Producing a reasonable summary is then considered to be evidence of an adequate understanding of the source text.

References

Cohen, Philip R. Jerry Morgan, and Martha E. Pollack (eds) 1990. Intentions in communications. Cambridge, MA; London, UK.: Bradford Book and MIT Press.

Grosz, Barbara J., Martha E. Pollack, and Candace L. Sidner 1989. Discourse. editor, Foundations of cognitive science, ed. by Michael I. Posner, 437–68. Cambridge, MA; London, UK: Bradford Book and MIT Press.

Grosz, Barbara J., and Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. Computational Linguistics 12(3), 175–204.

Kamp, Hans, and Uwe Reyle. 1993. From discourse to logic. Introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory, Vol.

42 of Studies in linguisitics and philosophy, 2 vols. Dordrect: Kluwer.

Kintsch, Walter, and Teun A. van Dijk. 1978. Toward a model of text comprehension and production. Psychological Review 85(5), 363–94.

Mann, William C., and Sandra A. Thompson. 1988. Rhetorical structure theory: toward a functional theory of text organization. Text 8(3). 243–81.

Morgan, Jerry L. and Georgia M. Green. 1980. Pragmatics and reading comprehension. Theoretical issues in reading comprehension. Perspectives from cognitive psychology, linguistics, artificial intelligence, and education, ed. by Rand J. Spiro, Bertram C. Bruce, and William F. Brewer, 113–40. Hillsdale, NJ: Lawrence Erlbaum.

Scha, Remko J. H., Bertram C. Bruce, and Livia Polanyi. 1987. Discourse understanding. Encyclopedia of artificial intelligence. Vol. 1, ed. by Stuart C. Shapiro, 233–45. New York: John Wiley.

Schank, Roger C., and Christopher K. Riesbeck. 1981. Inside computer understanding: five programs plus miniatures. Hillsdale, NJ: Lawrence Erlbaum.

Singer, Murray, 1990. Psychology of language. An introduction to sentence and discourse processes. Hillsdale, NJ: Lawrence Erlbaum.

van Dijk, Teun A. 1977. Semantic macro-structures and knowledge frames in discourse comprehension. Cognitive processes in discourse comprehension, ed. by Marcel A. Just and Patricia A. Carpenter, 3–32. Hillsdale, NJ: Lawrence Erlbaum.

Walker, Marilyn A., Aravind K. Joshi, and Ellen F. Prince (eds) 1998. Centering theory in discourse. Oxford, UK: Oxford University Press.

UDO HAHN

Thai and Tai Languages

Thai (Bangkok Thai, Standard Thai, Central Thai, or Siamese) is the national language of Thailand. It is spoken by approximately 60 million speakers, at least half of whom speak other languages and dialects as their mother tongues. The standard language, based on the speech of the elite and the upper middle class in Bangkok, is used as the medium of instruction in schools throughout the country. Regional dialects outside Bangkok and the Central Plains include Northern Thai (Kam Muang or Lanna), Northeastern Thai (Isan or Lao), and Southern Thai (Paktai). Although these varieties are generally considered dialects of Bangkok Thai, they are more closely related to other languages of the same family spoken in neighboring countries. For example, Northern Thai is more similar to Shan or Tai Yai in northern Burma, and Northeastern Thai to Lao. Strong regional identities exist in these regions, particularly among speakers of Southern Thai (most of whom are Muslim Malays) and northeasterners who call themselves and their language Lao. The name Thai-Isan is a newly coined term introduced by the Thai kingdom to promote nationalism among northeastern speakers. Apart from Thai and its dialects, other languages spoken in Thailand include Chinese (mainly in Bangkok and the central region), Malay in the south, and Mon-Khmer languages in the northeast. In addition, there are a number of hill tribe languages such as Karen, Hmong, Yao, Lisu, Lahu, and Akha in Tak, Chiang Mai, Chiang Rai, and Mae Hong Son in the north. Other Tai varieties are also spoken in scattered communities, such as Tai Yai in Mae Hong Son, Tai Ya and Tai Lue in Chiang Rai, and Phu Thai in Nakhon Phanom.

The word Tai (without the *h*) has two major uses: (1) it refers to a group of languages, including Thai and Lao, which are spoken in a vast geographical area of Southeast Asia ranging from southwestern China and northern Vietnam to the Indian state of Assam; (2) it is the name of a language family, the Tai family, a subgroup of Tai-Kadai or Kam-Tai. The term Kadai was coined by Paul K. Benedict to refer to a group of languages closely related to Tai such as Li (Hlai), Lakkia (Lajia), and Gelao, which are spoken mainly in China. The relationships between Tai and Kadai or Tai and Kam (Dong) languages as well as between Tai-Kadai and other language families in Southeast Asia are still controversial. According to Benedict, the Tai family is a branch of Austro-Tai, because it shows strong affinities with Austronesian, another major family in Southeast Asia comprising more than 1,000 languages. Scholars in China, on the other hand, view Tai (so-called Daic) as a branch of Sino-Tibetan, a superfamily that consists of Chinese (Sinitic), Tibeto-Burman, and Hmong-Mien (Miao-Yao). A view generally held by linguists today is that Tai-Kadai or Daic is a separate language family, which has no proved connection with either Austronesian or Sino-Tibetan. Hypotheses about Proto-Tai have been debated since the turn of the twentieth century and have motivated a good number of works on comparative Tai linguistics, particularly in the areas of pronunciation and vocabulary.

Within the Tai family, Fang Kuei-Li distinguished three branches: Southwestern Tai (comprising the well-documented languages of Assam, Burma, Yunnan, Thailand, Laos, and northern Vietnam), Central Tai (the most notable members are Nung and Tho spoken in northern Vietnam and areas of Guangxi adjacent to northeastern Vietnam), and Northern Tai (e.g. Northern Zhuang, Buyi, Yay, Saek, and Li). Zhuang is spoken in the Zhuang Autonomous Prefecture in Guangxi and Wenshan County in Yunnan. The Zhuang constitute the largest of China's 56 official minority nationalities, numbering nearly 20 million speakers. There are two main Zhuang languages: Northern Zhuang (already mentioned) and Southern Zhuang. The latter shows affinity with Central Tai languages and is classified as a member of this group. Spoken in the upper Red River area in Vietnam, Yay (Giy) has been richly described by William J. Gedney. Saek (previously classified as a Mon-Khmer language) is spoken in Laos and Thailand. It has been reported to maintain the archaic final consonant -*l.* Saek has also been investigated by Gedney.

In China, speakers of Tai languages are centered in autonomous prefectures and counties in the southwestern provinces of Yunnan, Guizhou, Guangxi, along the border region of Hunan, and Hainan Island. They are officially recognized as the Dai (China's Latin-script equivalent of *Tai,* referring to speakers of the southwestern branch), Zhuang, Buyi, Dong, Li, and Gelao. The majority of the Dai speak Tai Lue and are concentrated in Xishuangbanna Dai Autonomous Prefecture in Yunnan. Tai Nuea or Tai Luea is spoken in Dehong-Dai-Jinpo Autonomous Prefecture along the Yunnan–Burma border. This language shares several characteristics with Tai varieties in Shan of upper Burma. White Tai (Tai Don), Red Tai (Tai Daeng), and Black Tai (Tai Dam), all belonging to the southwestern branch, are spoken in various places in northern Vietnam. They are mutually intelligible and can be classified as dialects of a single language. Black Tai speakers, numbering approximately 500,000, were originally centered in Muang Thaeng (Dienbienphu) in Vietnam but they have now settled in Laos and several provinces in Thailand.

With a writing system based on Indic-derived scripts, several Tai languages have enjoyed long literary traditions. Although Ahom has become virtually extinct as a spoken language since the nineteenth century, its literary legacy provides an important clue to Tai-Ahom worldviews. The use of Lanna script in northern Thailand is today limited to religious texts. The use of the present-day Thai script for the standard written language in Thailand can be traced back to the reign of King Ramkhamhaeng during the Sukhothai period in the mid thirteenth century. The Sukhothai writing system was borrowed from Khmer

(Cambodian) and subsequently developed to accommodate special characteristics of the Thai language as it was spoken then. On the evidence of the Ramkhamhaeng Inscription (1292 CE), Sukhothai Thai included four kinds of syllables. The first three categories, ending in a vowel or sonorant (so-called *unchecked, live,* or *free* syllables), used contrasting tones (see below), whereas the fourth kind, ending in stops /p t k/ (*checked* or *dead* syllables), exhibited no tonal contrasts. In the Ayutthaya period (fourteenth to eighteenth centuries CE), the language underwent two major changes, which resulted from tonal splitting and from borrowing words from Khmer, Pali, and Sanskrit. The tonal split, which also affected other Tai languages, although at a much earlier date, was conditioned by the nature of the initial consonant and split the system of three tones into a system of five in Modern Thai. Some Tai languages have been reported to have as many as nine tones.

The inventories of sounds in Central and Northern Tai are generally different from those in the Southwestern branch. For example, Nung and Tho have peculiar initials such as *py-* or *phy-* for words that share ancestry with those beginning with *t-* in the Southwestern varieties. Li (Hlai), a Northern Tai variety from Hainan Island, has preglottalized stops and contains the rare voiced stop *g.* Southern Zhuang contains an aspirated stop series but Northern Zhuang has only unaspirated stops.

A syllable in Thai and Tai languages consists of an initial consonant or consonant cluster, a vowel or diphthong, and a tone, and may or may not have a final consonant. When the final consonant is present, it is either a voiceless stop /p t k ʔ/, a nasal /m n/, or a semivowel /y w/. The glottal stop /ʔ/ is an innovation in modern Tai languages; it probably did not exist in the ancestor language Proto-Tai. Words that contrast in tones convey different meanings. For example, the following words in Bangkok Thai mean different things depending on the pitch with which they are pronounced: *naa1* 'field', *naa2* 'sentence-final particle', *naa3* 'face', *naa4* 'mother's younger sibling', *naa5* 'thick'. The numbers 1–5 refer to different qualities of pitches: 1 = mid level, 2 = low level, 3 = falling, 4 = high level, and 5 = rising. Checked syllables have fewer tonal possibilities than unchecked syllables. Black Tai, for example, chooses from six tones on unchecked syllables but only two tones on checked syllables. Weak syllables are also common in Thai and Tai languages. The tone on such syllables is usually neutralized. The Thai *kra-* in *krathaʔ4* 'pan', *krathaŋ5* 'flowerpot', is a clear example of this type.

Most inherited Tai words are monosyllabic. The situation in Thai is more complicated due to words borrowed from other languages. Pali loans and compounds

greatly increase the number of polysyllabic words in the language and distinguish it from the Tai languages of China, which are heavily influenced by Chinese. Contacts with other languages thus result in finer lexical distinctions between dialects of a single language. For example, Tai Ya speakers from Muang Ya in Yunnan have difficulty communicating with their relatives in a transplanted community in Chiang Rai. Another common feature of Thai and Tai as well as non-Tai languages in Southeast Asia is the use of doublets and elaborate expressions. These are two- to four-syllable words which contain repeated or synonymous information. They usually convey idiomatic or figurative meanings and characterize playful speech. For example, the expression in Bangkok Thai *kin1 khaaw3 kin1 plaa1* (eat + rice + eat + fish) simply means 'to eat' but it is composed of two identical words (first and third syllables) and two words with similar meanings (second and fourth syllables).

Grammatical relations in Thai and Tai languages are indicated by word order, not by word inflections. Subject–Verb–Object is the usual pattern, although other patterns occur: Object–Subject–Verb (for object emphasis), Verb–Subject (stating existence), and Subject–Object–Verb (as in Khamti, a Tai variety in Ahom). An important characteristic of spoken and written texts is that virtually everything except verbs or adjectives can be omitted. Pronouns are especially prone to deletion. Nouns do not have plural forms, and verbs express no tense. To indicate whether actions have already taken place or will take place, the speaker counts on the addressee's interpretation and makes use of temporal adverbs, preverbal auxiliaries, as well as serial verb constructions (series of verbs that occur in concatenation without intervening words). It is still controversial as to whether there are true adjectives and prepositions in these languages, as the former generally behave in the same way as verbs and the latter can function as nouns.

Classifiers (category-words used in e.g. counting, somewhat like English *a slice of bread, a slice of pizza*) are found throughout the family, suggesting that this is an inherited trait from Proto-Tai. However, there are variations among the languages and dialects. For example, Bangkok Thai and Tai Lue have the Southeast Asian pattern with the following order: noun + numeral + classifier, whereas Black Tai and White Tai follow the Chinese pattern with the order numeral + classifier + noun. Classifiers can substitute for modified nouns in both Thai and less well-known varieties such as Tai Lue and Tai Nuea. These linguistic forms are generally used to indicate shape, size, flexibility, and animacy. Although classifiers are similar in grammar and meaning in various Tai languages, they differ in number and usage. There is a tendency for some

classifiers in Bangkok Thai to be of limited use and some to be used more extensively. For example, *chɯak3* 'classifier for elephants' is being replaced by *tua1* 'classifier for animals' in young people's speech. In the same way, *ʔan1* 'classifier for things' has been extended to cover other nouns that formerly required special classifiers.

Pronouns in Thai and Tai languages are of special interest to linguists, as there are many ways to express 'I' and 'you' depending on the sex of the speaker, the relationship between the speaker and the addressee, and the formality of the situation. Kinship terms, words expressing occupations, and nicknames can be used in this context. Furthermore, these languages employ a number of sentence-final particles, another areal feature in Southeast Asia. These particles are used to perform different kinds of speech acts and convey the speaker's attitudes toward the utterance. Like pronouns, they are used in accordance with the social status of the speaker with respect to the addressee, degrees of intimacy, and politeness. According to Gedney, pronouns and particles are what distinguish languages of this family which are far apart from one another, rendering them mutually unintelligible.

References

Benedict, Paul K. 1975. Austro-Thai language and culture, with a glossary of roots. New Haven, CT: Human Relations Area Files Press.

Bickner, Robert J., John Hartmann, Thomas J. Hudak, and Patcharin Peyasantiwong (eds.) 1989. Selected papers on comparative Tai studies, Ann Arbor: University of Michigan.

Compton, Carol J., and John F. Hartmann (eds.) 1992. Papers on Tai languages, linguistics, and literatures in honor of William J. Gedney on his 77th birthday. Monograph Series on Southeast Asia Occasional Paper No. 16. DeKalb: Center for Southeast Asian Studies, Northern Illinois University.

Diller, Anthony. 1994. Tai languages: varieties and subgroup terms. Thai-Yunnan Newsletter, No. 25, ed. by Gehan Wijeyewardene. Canberra: Research School of Pacific Studies, Australian National University.

Edmondson, Jerold A., and David S. Solnit (eds.) 1997. Comparative Kadai: the Tai branch. Publications in Linguistics No. 124. Dallas: Summer Institute of Linguistics and the University of Texas at Arlington.

Gedney, William J. 1991. The Yay language glossary, texts and translations, ed. by Thomas J. Hudak. Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan.

———. 1992. The Saek language glossary, texts and translations, ed. by Thomas J. Hudak. Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan.

Hartmann, John F. 1984. The linguistic and memory structure of Tai-Lue oral narratives. Monograph Series B-90. Canberra: Australian National University.

Hudak, Thomas J. 1987. Thai. The world's major languages, ed. by Bernard Comrie. London: Croom Helm, and New York: Oxford University Press.

Li, Fang-Kuei. 1977. A handbook of comparative Tai. Oceanic Linguistics Special Publications No. 15 Honolulu: University of Hawaii Press.

Terwiel, B.J., and Ranoo Wichasin (translators and editors) 1992. Thai ahoms and the stars: three ritual texts to ward off danger Ithaca: Cornell University Southeast Asia Program.

Wijeyewardene, Gehan. 1990. Thailand and the Tai versions of ethnic identity. Ethnic groups across national boundaries in mainland Southeast Asia. Singapore: Institute of Southeast Asian Studies.

KRISADAWAN HONGLADAROM

Thematic Structure

Thematic structure is a list of specifications in the lexicon about the thematic relations that hold between a predicate and its arguments. In some approaches, 'thematic structure' is referred to as 'thematic grid', 'theta grid', or 'argument structure'.

'Argument' and 'thematic relation' are lexical semantic notions. The event or state of affairs a verb denotes entails a fixed number of participants, which mostly have to be expressed in every well-formed sentence in which the verb is contained. These are the arguments of the verb. Although obligatory syntactic realization is not a necessary criterion for the identification of arguments, it is true that an argument is still implicit in the meaning of the verb even if it is not expressed in a sentence. In *Fred eats an apple with great relish*, for example, the expressions *Fred* and *an apple* realize the arguments of the verb *eat*. If *eat* is used intransitively, e.g. *Fred is eating with great relish*, we still understand that Fred is relishing something he eats. The lexical entry of a verb specifies for each argument the way in which it is involved in the denoted event. In the example above, Fred is the instigator of the eating event and the apple is the object that is affected. The roles participants may have in events denoted by verbs are the thematic relations. Other common names in different theories are 'thematic' or 'theta roles' ('Θ roles'), 'semantic roles', 'participant roles' and so on. They are generally subsumed under labels like 'agent', 'patient', 'goal', 'beneficiary', and 'experiencer'. Theories differ with respect to the number of labels they assume and the way in which they define the corresponding thematic relations.

In linguistic theories, thematic relations serve as an interface between lexical semantics and syntax. In different theories, there are various syntactic generalizations that refer to thematic relations, but almost every theory uses them overtly or covertly to predict the syntactic realization of arguments, e.g. as subject, object or indirect object, or with respect to case marking or syntactic alignment of the corresponding noun phrase.

The notion 'thematic relation' was first introduced in Gruber's *Lexical structures in syntax and semantics* (1976) and in Fillmore's *The case for case* (1968), who called them 'deep cases'. Fillmore assumed that each argument is associated with a preposition that corresponds to its deep case. Every argument is inserted with its preposition into a unique syntactic position in deep structure. In the examples in (1), *John* has the deep case 'agentive', *the door* the deep case 'objective', and *the wind* the deep case 'instrumental'. The agentive and the instrumental take the prepositions *by* and *with*, respectively, cf. (1a), (1c), and the preposition of the objective is zero, cf. (1a–d).

(1a) John opened the door with the key.
(1b) The key opened the door.
(1c) The door was opened by John.
(1d) The door opened.

'Agentive' is, in Fillmore's theory, 'the case of the typically animate perceived instigator of the action identified by the verb'; 'instrumental' is 'the case of the inanimate object or force causally involved in the action or state identified by the verb', and 'objective' 'should be limited to things that are affected by the action or state identified by the verb'; cf. Fillmore (1968:24f.).

In the course of the derivation of an active sentence, an agentive moves to subject position, where its preposition is deleted, cf. (1a). If no agentive is present at deep structure, the instrumental will move to subject position where its preposition is deleted. Even objectives can move to subject position, if they are the only argument in active sentences, cf. (1d), or in passive sentences, then the agentive remains with its preposition in its original deep structure position, cf. (1c).

Later approaches in Generative Grammar departed from the postulation that arguments are inserted into deep structure with prepositions according to their thematic relations. They nonetheless continued to adhere to the idea that there is a one-to-one correspondence

between thematic relations and syntactic positions. Chomsky (1981) introduced in his *Lectures on government and binding* (GB) the 'theta criterion', which poses the following twofold restrictions on thematic structure: (1) each argument is assigned one and only one theta role; and (2) each theta role is assigned to one and only one argument. In (1a) and (1b), we have two different variants of the verb *open.* In (1a) the agent role is assigned to the subject, the direct object is assigned the theme role, and the prepositional phrase is an instrument(al). (1b) differs from (1a) in that the subject bears the instrument role. Both sentences meet the theta criterion because there is no argument without a theta role and no theta role is assigned twice in one sentence. The theta criterion correctly rules out sentences like *The key opened the door with the bit,* where the instrument role is assigned to *the key* and *the bit* (unless the subject is interpreted as a character in a fable or cartoon), or *John and the key opened the door,* where the collective subject is assigned the instrument and the agent role at the same time. In GB it was assumed that the verb in (1d) assigns the theme role to the direct object, but that it fails to assign case to this argument. It was postulated that a verb's failure to assign case to the direct object always correlates with its failure to assign a theta role to its subject. In other words, in the underlying structure of (1d) the subject position is empty and the direct object lacks a case feature. In order to receive case, the phrase moves to the subject position and thus yields the surface structure in (1d). Passive sentences like (1d) were also accounted for by movement of the direct object to the subject position, but in this case the theta role of the subject and the case of the direct object was assumed to be absorbed by the passive verb. (The apparent agent *by John* in (1d) does not represent an argument of the passive verb. Instead it is assumed to be an adjunct.)

While in GB for most arguments a thematic and a grammatical relation is specified in the lexical entry of the verb, other theories provide algorithms that predict the grammatical relation or the morphosyntactic realization of arguments from their thematic relations. Such theories are called 'linking theories'. The examples in (1) illustrate the fact that, in English, agents, instruments, and themes or patients can all become subject of active sentences. However, an agent will always be realized as subject while an instrument cannot become subject if the thematic structure contains an agent, and a patient or a theme can only become subject if it is the only argument in the thematic structure. These generalizations seem to hold cross-linguistically for languages that have accusative case systems like English. In some contemporary approaches, the different ranking of thematic relations with respect to

Figure 1

the choice of subject and the application of other syntactic rules is reflected by thematic hierarchies, cf. Van Valin's (1993) Role and Reference Grammar (RRG: 41) (see Figure 1).

On this continuum, thematic relations that are nearer to the left end share more properties with the agent ('the willful, volitional, instigating participant', cf. RRG: 42), while thematic relations that are nearer to the right end are more patient-like ('the nonwillful, noninstigating, maximally affected participant', cf. RRG: 42). 'Effector' is a generic term for instruments and forces, experiencers are, for example, nonwillful instigators of perceiving events, and themes undergo a change of location but not a change of state or condition. Sources, paths, goals, and recipients are grouped together, because they do not compete with each other for subjecthood or other syntactic relations. They do, however, compete with the theme. The mapping of thematic relations to syntactic relations proceeds in the following way: (1) The highest ranking thematic relation in a thematic structure is assigned the macrorole 'actor' and the lowest ranking thematic relation is assigned the macrorole 'undergoer'. If a verb has more than one argument, in accusative languages the actor is mapped to the subject and the undergoer to the direct object. Any remaining arguments are assigned cases or prepositions according to their thematic relations. The choice of the particular case or preposition is language-specific. While in example (1a) the agent *John* outranks the theme *the door*, in (1b) it is the effector *the key*. Thus, *John* and *the key* are actors in (1a) and (1b), respectively, and *the door* is undergoer in both sentences. In (1c) and (1d) *open* has only one argument, the theme *the door*. Since the verb does not express an activity in either case, the argument is an undergoer, but this does not affect the choice of the grammatical relations, since only arguments are always realized as subjects.

One advantage of RRG is that it captures distinctions between thematic relations like agent and effector and at the same time allows for generalizations over a wide range of thematic relations, for example, with respect to linking. Still a disadvantage of RRG is that it operates with role labels. A thematic relation that is covered by a role label is defined by a number of related properties (e.g. Van Valin's definition of the agent above), which may not all apply to every argument for which the role label was intended. Because in every theory there is a residue of arguments that do not

fit in any role label category or that do fit equally in two or more categories and whose assignment to one or another label often seems to be subject to theoretic bias, many critics have rejected the theoretic value of thematic relations altogether. Dowty's (1991) *Thematic proto roles and argument selection* (TPR) offers a solution to this problem: in this model, thematic relations are characterized by already familiar properties like volitionality, etc. that are entailed by the verb of its arguments. These properties come from two finite sets that define the prototypical agent and the prototypical patient, respectively, and combine freely with each other. The proto-agent properties are (1) volitional involvement in the event or state, (2) sentience/perception, (3) causing an event or state in another participant, and (4) movement relative to another participant. The proto-patient properties are (1) undergoes a change of state, (2) incremental theme (a participant that successively undergoes a change of state in an event that does not end before every part of the participant has been affected, e.g. *a pizza* in *Bert ate a pizza*), (3) causally affected by another participant, and (4) stationary relative to movement of another participant (TPR: 572). Dowty's model permits a very fine-grained distinction of thematic relations since arguments can be characterized by a great variety of properties from one or both concepts, or even by the absence of proto-role properties. On the other hand, due to the limited number of proto-role properties and the mutual independence of the properties, it is possible to generalize over very large classes of arguments that share a certain property, regardless of other possible proto-role properties these arguments may have. The linking rules in TPR correctly predict that the only argument or the argument that has the greatest number of proto-agent properties in the thematic structure of a verb is lexicalized as the subject. The argument that bears the greatest number of proto-patient properties becomes direct object. For thematic structures with three arguments, it holds that the argument having less proto-agent properties than the subject and less proto-patient properties than the direct object will be the indirect object.

Because of its obvious advantages, Dowty's model has received wide acknowledgment in the literature. However, his linking rules have been subject to the criticism that they refer to grammatical relations and not to morphosyntactic features. As Primus's (1999) *Cases and thematic roles* exhaustively shows, notions like subject and object only apply consistently to languages with accusative case systems, but not to ergative and split intransitivity languages. As a consequence, Primus modified Dowty's model in order to achieve universal validity, like it was intended by Dowty in the first place.

References

Chomsky, Noam. 1981. Lectures on government and binding. Dordrecht, The Netherlands: Foris.

Dowty, David. 1991. Thematic proto-roles and argument selection. Language 67.

Fillmore, James. 1968. The case for case. Universals in linguistic theory, ed. by Emmon Bach and Robert Thomas Harms. New York: Holt, Rinehart and Winston.

Gruber, Jeffrey. 1976. Lexical structures in syntax and semantics. Amsterdam, The Netherlands: North-Holland Publications.

Jackendoff, Ray. 1983. Semantics and cognition. Cambridge, MA: MIT Press.

———. 1987. The status of thematic relations in linguistic theory. Linguistic Inquiry 18.

Primus, Beatrice. 1999. Cases and thematic roles: ergative, accusative and active. Tübingen, Germany: Niemeyer.

Van Valin, Robert. 1993. Advances in role and reference grammar. Amsterdam, The Netherlands and Philadelphia, PA: Benjamins.

KERSTIN BLUME

See also **Chomsky, Noam; Fillmore, Charles; Generative Grammar; Grammatical Function; Lexicon: Overview.**

Tibetan

Tibetan belongs to the Bodish branch of the Tibeto-Burman division of the Sino-Tibetan language family. It is closely related to Newar and Tamang, indigenous languages of Nepal; distantly related to other Tibeto-Burman languages such as Akha, Lahu, Karen, and Burmese; and remotely related to Chinese. It is spoken by approximately five million people in six different nation-states: China, Burma, India, Nepal, Bhutan, and Pakistan. Only in Bhutan is it the national language. Significant groups of Tibetan speakers are also found in exile communities in India, Nepal, Bhutan, and several western countries: at least 100,000 Tibetans fled

Tibet in 1959 when China took it over. Of the five million Tibetans, 2,100,000 speakers are concentrated in the Tibetan Autonomous Region, with Lhasa as the main municipality, and 2,500,000 speakers are found in several Tibetan autonomous prefectures and counties in four western provinces of China: Qinghai, southwest Gansu, western Sichuan, and northwest Yunnan. The remainder are in the Himalayan regions covering the northern tip of Burma, northern Nepal, Bhutan, the Indian states of Arunachal, Sikkim, Himachal Pradesh, and Uttar Pradesh, the states of Jammu and Kashmir (divided between India and Pakistan), and a Pakistani district of Baltistan.

Tibetan-speaking people in the Himalayas are known as *Bhotia* (from the word *bod,* the name of Tibet). A minority group in eastern Nepal numbering approximately 14,000, more commonly known as the Sharpa (Sherpa), speaks the southern variety of Central Tibetan. The Bhutanese call their national language Dzhongkha and do not classify themselves as Tibetans. Ethnic Tibetans in China are officially designated as the Tibetan nationality (*zàngzú* in Chinese or *bod-rig,* a newly coined term in Tibetan). The Jiarong (Rgyarong), Baima, and other ethnic groups of western Sichuan consider themselves Tibetans and are officially included within the Tibetan nationality, although they speak non-Tibetan languages as their mother tongues. The variety of Tibetan spoken in Baltistan is called Balti, but it acquires another name (Purik) when spoken on the Indian side.

With historical evidence dating back to the seventh century, Tibet was an important empire of Central Asia extending its influence to the Tang dynasty of Imperial China and ruling over small kingoms such as the Nan Chao Kingdom (present-day Dali) in southwest China. It was in contact with several nation-states of Asia, including India, from where it borrowed a writing system. The Tibetan script—with 30 consonant and four vowel signs and a few punctuation marks—was modeled after the Brahmi script during the reign of King Srongtsan Gampo (617–650 CE). The language from this period through the ninth century, known as Old Tibetan, was used mainly to record royal documents and historical chronicles. A large number of Old Tibetan manuscripts found in the Dunhuang caves are among the most important materials for the study of the proto-history and proto-language of Tibet. The majority of these materials have been cataloged and preserved in national libraries in France and England.

Classical Tibetan (sometimes called *Chöke* by western scholars) has developed since the eleventh century. It remains in use as the medium for religious texts and other kinds of learned discourse. This literary variety was adopted as the liturgical language by Mongols and other non-Tibetan-speaking groups in Nepal who are influenced by Tibetan Buddhism such as the Tamang, Gurung, Thakali, and Manang. It has served as the standard written language for all learned documents in Bhutan. The written variety from the nineteenth century, which was developed from Classical Tibetan with influence from the Lhasa colloquial, is called Modern Literary Tibetan or Newspaper Tibetan. It is used widely in Tibetan communities in China and elsewhere as the standard medium for newspapers, magazines, and other kinds of modern writings, including radio broadcasts. Other varieties that have developed from Classical Tibetan but are based on regional idioms also emerged, but their use is limited to particular regions. Written Tibetan, specifically Classical Tibetan, still represents the pronunciation of the language as it was in the ninth century, when it underwent a major reform. Hence, there are considerable differences between Written Tibetan and modern spoken dialects. For example, the greeting expression *bkra-shis bde-legs* 'Good luck' is pronounced *trashi tele* in Lhasa speech.

Spoken Tibetan is divided into several dialects and subdialects that are quite different from one another. The number of dialects varies, depending on the geographical areas under investigation. Most authorities agree that there are at least four major dialects: Central Tibetan (Ü-tsang), Northeastern Tibetan (more commonly known as Amdo), Eastern Tibetan (Kham), and Western Tibetan (Töö). Central Tibetan, which includes Lhasa, the standard dialect, Shigatse, and other subdialects, is spoken in the Tibet Autonomous Region (TAR) and along the Tibet–Nepal border. Amdo is spoken in Gansu and Qinghai provinces. Kham is found in Sichuan, Qinghai and Yunnan provinces, and some parts of TAR. Western Tibetan includes Ngari in TAR, Ladakhi in India, and Balti in Pakistan. Chinese linguists pay attention only to the first three groups, as they are the major dialects spoken in China.

Within Kham Tibetan, two groups are distinguished: valley or sedentary dialects (*Rongke*) and nomadic dialects (*Drokke*). Each category is further divided into several subdialects. Amdo too is divided into valley dialects and nomadic dialects, although the difference between these two categories is not so great as in Kham. There is a continuum of mutual intelligibility among the speakers of these dialects. Amdo speakers are reported to have problems understanding Lhasa speakers and vice versa, and thus it is common for the two groups to resort to Chinese as a lingua franca. For each region, one subdialect associated with cultural heritage, education, or political power is generally considered to be more prestigious than the others and serves as a regional lingua franca. Lhasa Tibetan is a subdialect of Central Tibetan, but because it is spoken

in the metropolitan and pilgrimage center and former seat of government, it is generally regarded by Tibetans as the most prestigious form. Lhasa Koiné, with loanwords from Nepali, Hindi, or English, is spoken in exile communities by refugees who came from various linguistic backgrounds. Other regional koinés in Tibetan territories in China include the Labrang dialect of Amdo and the Derge dialect of Kham.

The majority of words in Tibetan are disyllabic. A basic monosyllabic word consists of an initial and a vowel and may or may not have a final consonant, e.g. *bod* 'Tibet' (*phöö* in the Lhasa dialect; *wot* in the Amdo dialect). Consonant clusters are common in written Tibetan and archaic dialects—those containing archaic features and therefore closer to written Tibetan—namely Amdo and western dialects. The number of initial consonants varies from one dialect to another. There are 28 consonants in Lhasa Tibetan. Kham valley dialects have about 40 consonants. Tibetan presents an interesting case for the study of the origin of tones (i.e. intonational differences that distinguish word meanings), as this innovative feature of distinguishing meanings by musical pitch alone is found only in some modern dialects. Central Tibetan and Kham Tibetan exhibit a well-developed system of lexical tones, whereas Amdo and Western Tibetan do not have tones. The number of tones ranges from two to four. Tibetan dialects in Nepal display a two-tone system (high tone vs. low tone). Most Kham dialects contain four tones (high tone, low tone, rising tone, and falling tone). Initial and final consonants are relatively simple in these tonal dialects. In addition to tones, some of these dialects also developed a rich inventory of diphthongs.

Like the majority of Tibeto-Burman languages, Tibetan has the subject–object–verb word order. Grammatical relations are expressed by means of case particles. Five cases are distinguished for most spoken and written varieties: ergative–instrumental, ablative, absolutive, genitive, and dative–locative. The genitive case particle also functions as a relative clause marker. The ergative and the instrumental case markers are homophonous but they are attached to nouns with different semantic roles: the ergative particle marks an actor, which is chiefly animate, whereas the instrumental particle marks an instrument. Likewise, the dative and the locative case particles are identical in form, but they have different functions: the former marks a recipient or a patient (an entity, generally an animate being, affected by the action); the latter is attached to a place or time. Written Tibetan is a prototypical example of an ergative language: subjects of verbs that can occur with objects require ergative case marking, whereas objects of such verbs and subjects of verbs that never take objects behave differently: they are unmarked, which

indicates absolutive case. Lhasa Tibetan is often cited by linguists as a language with a 'split' ergative-marking pattern: ergative subjects are required only when reported actions have already taken place. Therefore, the subject in *khōŋ-ki tà tçí sὲε sōŋ* 'He killed a tiger' requires an ergative case particle because the verb indicates a past event, whereas the subject of *khōŋ tà tçí sὲε-ki rè* ''He will kill a tiger' does not.

Tibetan nouns express neither gender nor number. Plurality is indicated by the plural marker *tsho* as well as *dag* and *rnams* (the latter two are more common in written texts). In modern dialects, these plural particles are generally employed only with pronouns. Adjectives follow nouns. Determiners (words that modify noun phrases) include the number 'one' *gcig,* which functions as an indefinite article, and the spatial demonstratives *di* 'this' and *de* 'that' functioning as definite articles. Negation is indicated by *mi-* or *ma-,* which is prefixed to the main verb. In Lhasa Tibetan, it is more common to use sentence-final negative verbs, i.e. *min* and *me,* when the subject is first person 'I', and the negative prefixes when the subject is second or third person, i.e. 'you' or 'he or she or it'. A sentence may consist of a single clause or a series of clauses chained together with the main clause as the last element. Because of this characteristic, Tibetan is typologically classified as a 'clause-chaining' language.

Verbs in Old Tibetan and Classical Tibetan, most of which are monosyllabic, have variant forms corresponding to tenses (future, present, past) and mood (imperative). In Central Tibetan and Kham Tibetan, this distinction is neutralized: there is only one form for each verb. Instead, sentence-final auxiliary verbs are employed to convey, for example, tense. These auxiliary verbs and verbs equivalent to 'to be' or 'to exist' are used according to person (whether or not the subject is first person) and evidentiality (whether or not the person has witnessed the event described in the utterance). Person marking in Tibetan is different from person agreement, a common feature of languages of the Himalayan branch of the Tibeto-Burman group. The emergence of person and evidential markings in modern spoken dialects of Tibetan has attracted a great deal of attention from linguists. These phenomena are also found in other Bodish languages.

Tibetan has a well-defined and productive system of honorifics (*zhesa*). These are special words (mainly nouns and verbs) used when referring to Buddhas and deities and in everyday conversation when one talks with people of higher social standing, elderly and respectable people including monks, officials, teachers, and one's own parents and elder siblings. Honorific vocabulary, which stands in parallel with ordinary vocabulary (*phal-skad*), is common in Old Tibetan texts and has developed into a complicated

system in Lhasa Tibetan. It is found in most of the dialects but in a relatively less sophisticated manner. Contrary to stereotypes among native speakers, honorifics are used among Kham speakers, but mainly when they engage in conversation with monks, high officials, or the elderly. The following expressions represent two ways of saying 'his hat', with the structure [he-genitive hat]: *khō-ki zhamo* (ordinary speech) and *khōŋ-ki üzha* (honorific speech). The honorific compound *üzha*, which consists of *ü*, the honorific root for 'head,' and *zha*, the ordinary root for 'hat,' is typical of how honorific nouns are formed in Tibetan.

Given that most Tibetan-speaking areas are inaccessible for fieldwork, both a basic description and an investigation of contemporary aspects of Tibetan languages and dialects still await linguists today. With continuing sociocultural and political changes, Tibetan remains one of the most challenging areas of study for the twenty-first century.

References

Beckwith, Christopher, I. 1992. Deictic class marking in Tibetan and Burmese. Proceedings of the first annual meeting of the Southeast Asian Linguistic Society, ed. by Martha Ratliff and Eric Shiller. Tempe: Arizona State University Program for Southeast Asian Studies.

Beyer, Stephen. 1992. The classical Tibetan language. Albany, NY: State University of New York Press.

Bielmeier, Roland. 1985. A Survey of the development of Western and Southwestern Tibetan dialects. Soundings in Tibetan civilization: Proceedings of the 1982 seminar of the International Association for Tibetan Studies held at Columbia University, ed. by Barbara Nimri Aziz and Matthew Kapstein. New Delhi: Manohar.

Bradley, David. 1996. Tibetan. Atlas of languages of intercultural communication in the Pacific, Asia, and the Americas, ed. by Stephen A. Wurm, Peter Mühlhäusler and Darrell T. Tryon, Vol. 2. Berlin: Mouton de Gruyter.

DeLancey, Scott. 1998. Semantic categorization in Tibetan honorific nouns. Anthropological Linguistics 40. 1.

Denwood, Philip. 1999. Tibetan. Amsterdam and Philadelphia: John Benjamins.

Goldstein, Melvyn C., Gelek Rinpoche, and Lobsang Phuntsog. 1991. Essentials of modern literary Tibetan: a reading course and reference grammar. New Delhi: Munshiram Manoharlal Publishers.

Makley, Charlene, Keith Dede, Hua Kan, and Wang Qingshan. 1999. The Amdo dialect of Labrang. Linguistics of the Tibeto-Burman Area 22. 1.

Tournadre, Nicolas. 1998. Manuel de tibétain standard. Paris: Asiatheque.

van Driem, George. 1998. Dzongkha. Leiden: Leiden University.

KRISADAWAN HONGLADAROM

Time and Tense

The tense of a sentence is a grammatical category that specifies a time at which the situation described by the sentence is the case. In a sentence like *I was often in Venice,* the past tense *was* indicates that the situation was the case before the time at which the sentence was uttered. Thus, tense mostly gives us information about the distance between a certain situation and the utterance time. Besides the grammatical category of tense, there are also lexical means of expressing location in time: temporal adverbs (*last year, yesterday,* etc., cf. *John left home **yesterday**),* prepositions of time *(in, at, on,* etc., cf. *John left home **at** lunch time),* and temporal conjunctions *(when, after, before* etc., cf. *John left home **when** the phone rang).*

Although tense is marked on the verb in many languages, many researchers assume that tenses modify sentences as a whole. Thus, a common analysis of *I was often in Venice* is PRET *(I often be in Venice),* where PRET is the abstract marker for preterite tense (simple past). Sentences in the present are analyzed in the same manner: PRES *(John go to the cinema)* is the analysis of *John goes to the cinema.* However, the treatment of the complex tenses like perfect and pluperfect is not that straightforward. Some analyze a complex tense as a single modifier. Thus, *John has gone to the zoo* would be represented as PERF *(John go to the zoo),* whereas *John had gone to the zoo* would be PLUPERF *(John go to the zoo).* Others assume that the complex tenses are combinations of modifiers, where the perfect combines PRES and PERF, and the pluperfect PRET and PERF. In these combinatorial accounts, *John has gone to the zoo* is analyzed as PRES *(PERF (John go to the zoo)),* whereas *John had gone to the zoo* would be PRET *(PERF (John go to the zoo)).* At first glance, the combinatorial account seems more attractive because it mirrors the actual words being used: present tense (*has*) and past tense (*had*) combine with perfect (*gone*). However, not all semanticists are convinced that tense on an abstract level necessarily keeps these modifiers separate. The theory of tense outlined here was heavily influenced by the work of Hans Reichenbach.

Another very influential source for tense theory was tense logic. The treatment of tenses as sentence modifiers, which was sketched above, goes back to tense logic. Tense logicians assigned indefinite meanings to the tenses. Compare the sentence *I turned off the stove,* represented as PRET *(I turn off the stove)*. This is considered true if and only if there is an indefinite time before the time of the utterance at which 'I' turn off the stove. But this indefinite meaning of the preterite is counterintuitive: anyone uttering the sentence will have a certain (definite) time of turning off in mind (for instance the time after the meal was ready). The point becomes even clearer with negation: *I didn't turn off the stove* does not mean that there is no turning off in the past. Instead, it means that there is a definite time in the past, and at this time 'I' did not turn off the stove. To say that tenses are definite means that a context is required for the listener to determine which exact point or stretch in time is being talked about. Relevant contexts are the immediate situation of utterance (in the case of the present tense) and times mentioned in previous discourse (in the case of the past tense).

Whereas 'classical' tense logic confined itself to single, isolated sentences without context, more recent work on tense has also focused on tense in discourse. Discourse may be 'temporally connected' or 'temporally free'. A sequence of two sentences forms a temporally connected discourse if the second sentence is temporally interpreted by taking elements from the first one. An example is: *Al went to New York. The others were there, too.* In contrast, temporally free discourse lacks a direct temporal connection. Compare: *Al went to New York. The others were there **once,** too.*

The division between sentences in the preterite whose events overlap each other and sentences where the events follow each other is related in a very systematic way to the types of the events being described. If both events are accomplishments or achievements, then the events are understood as happening in succession: *Jameson entered the room. He switched on the light.* If both sentences contain states or activities, then the events overlap each other: *Jameson cleaned the room. It was very cold in there.* However, if one of the two sentences contains a state or an activity, and the other an accomplishment or achievement, both overlap and succession are possible. The following is an example of succession: *He switched off the light. Now it was pitch-dark around him.* The darkness is due to the switching off and thus cannot overlap it. The following is an example of overlap: *He switched off the light. It was very cold in the room.* Interestingly, the progressive tense (marked with *-ing* in English) behaves like states or activities: if one of the two sentences contains a progressive, and the other an accomplishment or achievement, both overlap and succession is possible: *He awoke to the sound of her screeching. She was shaking him* (overlap); *He opened the door again. The man outside was smiling at him* (succession).

References

Benthem, J.F. van. 1983. The logic of time. Dordrecht: Reidel.

Comrie, B. 1985. Tense. Cambridge: Cambridge University Press.

Hinrichs, E. 1986. Temporal anaphora in discourses of English. Linguistics and Philosophy 9. 63–82.

Kamp, H., and U. Reyle. 1993. From discourse to logic. Dordrecht: Kluwer.

Klein, W. 1994. Time in language. London, New York: Routledge.

Nerbonne, J. 1985. German temporal semantics: three-dimensional tense logic and a GPSG fragment. Ann Arbor, MI: University Microfilms.

Ogihara, T. 1996. Tense, attitudes, and scope. Dordrecht: Kluwer.

Partee, B. 1973. Some structural analogies between tenses and pronouns in English. The Journal of Philosophy 70. 601–9.

Prior, A. 1967. Past, present and future. London: Oxford University Press.

Stechow, A. von. 1991. Intensionale Semantik—Eingeführt anhand der Temporalität. Konstanz: Arbeitspapier Nr.40, Universität Konstanz.

Stechow, A. von. 1995. On the proper treatment of tense. Proceedings of SALT V, ed. by T. Galloway and M. Simons.

MONIKA RATHERT

Tiv and Tivoid Languages

Tiv (Mitshi, Munshi) is spoken in Eastern Nigeria in an area all along the Katsina Ala and its tributaries, crossing the Benue where the Katsina Ala joins it and north of Makurdi, and in a few isolated villages across the border in Cameroon. Tiv's neighbors to the east are the Jukun; to the north Alago is spoken; to the east Eloyi, Agatu, and Igede; and to the south one finds a number of related languages/dialects, all of which are listed below. In the standard reference works it is classified as Bantoid—Benue-Kwa—Niger-Congo. Still, there is well-justified, long-standing discussion on just how close Tiv is to Bantu itself. Munshi is a derogative name for the Tiv

used by their neighbors. The term is a Hausa corruption of the Jukun term for the Tiv 'Mbitse' or 'Mbiche'.

Within Tiv proper very little dialect variation is known, even though it might be possible, based on differences in the realization of front vowels, to speak of the Iharev, Kparev, Kunav, Shitire, and Utisha dialects. The Shitire 'dialect' is used in most of the vernacular published literature.

To the Tivoid subgroup of Bantoid belong a series of very closely, potentially mutually intelligible languages spoken on either side of the Nigerian (N)–Cameroonian (C) border along the 6th latitude north. Included in this group are the following:

Abon (N), Avande (C), Batomo (C), Batu cluster (N), Bitare (N), Evant (N), Iceve-Maci (N), Iyive (N), Mesaka (C), Ndii (C, N) Oliti (C), Otank (N), Tiv (N), Tivi (C), and Yiive (C).

While there are around 2,500,000 speakers of Tiv, most of the other members of this subgroup are spoken by only a few thousand speakers.

Tiv was first documented by Clarke 1848—some number names and basic nouns—under the name 'Appa' and by Koelle 1854, who gave a lengthy word list. Significant linguistic works include the dictionary and grammar of Abraham and Malherbe, and the sketches of Lukas, Arnott, Sibomana, and Jockers. The language is used in primary schools and there are some readers for adult education. There is a Bible (1964) and a modest amount of liturgical literature. Tiv is used in regional radio broadcasts; newspapers or journals in Tiv existed until the early 1980s, but have since been discontinued.

Bohannan and Bohannan (1953:11) report that some people as far away as Bamenda, Cameroon used Tiv as a trade language. Today this no longer seems to be the case. Tiv does, however, enjoy a rapid expansion to the north, much to the dismay of its neighbors.

While Tiv is relatively well documented, for the other Tivoid languages the documentation is restricted to a few short word lists.

Some Structural Facts

Phonology. There are 26 consonants, six vowels, and three tones, low, mid, and high [ˋ ˉ ˊ] (Table 1). Tones are normally not marked in the official orthography. A number of morphophonemic rules are to be observed. Particularly interesting are the palatalization and labialization of stem-initial consonants under the influence of prefix vowels (see below) and stem alternations, in verbs due to the conjugational type and in nouns due to noun class prefixes.

Morphology, nouns. There are 12 noun classes grouped into 16 genders: 1|2; 1|4; 1|6; 5|4; 5|6; 5|6; 7|6; 7|6.a; 7|8; 9|4; 9|6; 14|8; 15|4; 15|6; 15|6.a; 6.a. Morphological agreement is shown in all categories of modifiers, in the pronominal system, and in the form of subject agreement on the verb. Noun class prefix as well as suffix markings are to be observed on the nouns and the adjectives. In Table 2, samples of each noun class each with one modifier category are given. The numbers to the left (and those given for the genders above) refer to the corresponding numbers of the Benue-Congo noun class system.

While only class (6a) nouns show clear prefixes and suffixes, e.g. ḿ-kúlé-m 'oil', remnants of class prefixes and/or suffixes can be established for all nouns. In some cases, the noun class prefixes have been incorporated into the stem: historical *u- is realized as labialization on the first consonant, e.g. kwāsé 'woman' compare kàsév 'women', while prefix i- causes palatalization of the first consonant, e.g. ínyáhá 'paddles', cf. náhá 'paddle'.

Modifiers always follow the head noun.

TABLE 1 Consonants and Vowels

	Bilabial	Labio-dental	Alveolar	Postalveolar	Palatal	Labio-velar	Velar	Glottal
Nasal	m		n		ny [ɲ]		ng [ŋ]	
Plosive	p, b		t, d			kp, gb	k, g	
Affricative			ts, dz	c [tʃ], j [dʒ]				
Fricative		f, v	s, z	sh [ʃ]			gh [γ]	
Lateral			l~r					
Approximant	w				y			h

	Front		Back
High	i		u
	e		o
			ọ[ɔ]
Low		a	

TABLE 2

	Sample Noun	Meaning	my	this, these	one	two	bad	Pro	subject concord	Rel
1	kwāsé	*woman*	wàmˊ	ngù	mòmˊ		ù bó	únˋ	á	ù
2	kàsév	*women*	ávˉ	mbáˋ		ú hálˉ	mbà bóvˊ	vé	vé	mbà
3	náhá	*paddle*	wámˊ	ngúˋ	mómˉ		ú bō	ú	ú	úū
4	ínyáhá	*paddles*	yámˊ	ngî̀		í hyálˉ	í bō	í	í	ī̀
5	íkyésé	*basket*	yámˊ	ngî̀	í mómˉ		í bō	í	í	ī̀
6	ákésé	*baskets*	ámˊ	ngáˋ		á hálˉ	á bō	á	á	áā
6a	ḿkúlém	*oil*	ámˉ	máˋ		ḿ hálˉ	mà bómˊ	má	má	mà
7	íny'álégh'	*money*	yághˉ	kî̀	í mómˉ		kì bóghˊ	kí	kí	kì
8	íkyáv'	*loads*	yávˉ	mbî̀		í hyálˉ	mbì bóvˊ	mbí	mbí	mbì
9	ìwyá	*dog*	yàmˊ	ngì	ì mòmˊ		ì bó	ì	ì	ì
14	ángév'	*disease*	ávˉ	mbúù	mómˉ		mbù bóvˊ	mbú	mbú	mbù
15	gbógh'	*stick*	ághˉ	kúˋ	mómˉ		kù bóghˊ	kú	kú	kù

Verbs. Verb stems (and also some nouns) may show internal alternations:

tèmà	tùmà	*sit*
nyímè	nyúmè	*refuse*
sèngè	sòngu	*slaughter*

Since not all verbs undergo these alternations, there is no immediate explanation for these changes although they appear to be remnants of an earlier (ATR) vowel harmony system.

There are traces of verb extensional suffixes. Five morphemes have been identified. It is not clear to what extent these processes are productive; they do not seem to play a role in marking participants.

sV	yóò	*announce*	yóò-sò	*talk a great deal*
rˋ	búghˋ	*open*	búghù-rˋ	*burst open*
ghˋ	pérˋ	*cross (a river)*	pérè-ghˋ	*cross over (road)*
mˋ	lèghˋ	*soft*	lèghè-mˋ	*be soft*
nˋ	pùù	*despise*	pùù-nˋ	*find fault with*

The tense–aspect–modality (TAM) system is extremely complex and only superficially described. A selection of the many forms that exist for each verb is given below. It should be noted that TAM is frequently marked with tones as well as TAM markers before the verb and vowel alternations after the stem.

	vèndà	*Refuse*
INFINITIVE	ù-vèndá-ń	*to refuse*
GERUND	vèndà-ń	*'refusing'*
RESTRICTED PAST PERFECT	ḿ-ˋvèndé	*I refused*
PRESENT HABITUAL I	mbá vé ˋvendá	*they usually refuse*
PRESENT HABITUAL II	Dàm ˋ ˋvèndé	*Dam usually refuses*
PRESENT HABITUAL III	mbá ˋvé véndá-ń	*they ususally refuse*
IMPERATIVE	ˋvèndá	*refuse!*
NEGATIVE IMPERATIVE	dé ˋvèndá-ń gá	*don't refuse!*
SUBJUNCTIVE, HORTATIVE	à vénda	*let him refuse!*

| PAST PERFECT | á vénde | *he refused* |
| *etc.* | | |

Syntax. Tiv is a basic SVO language. Sample sentences taken from the Tiv dialect spoken just across the Cameroonian border giving core participants show a strict S V, S V O, S V DAT O order:

Wán	lá	à-kpé		
Child	DEF	it-died		

The child died.

Wán	lá	à-yá	àyàbà	
Child	DEF	it-eat	plantain	

The child ate plantain.

Wán	lá	à-ná	tùɔ TM	àyàbà
Child	DEF	it-give	chief	plantain

The child gave the chief plantain.

Additional participants are expressed by a series of prepositions:

Wán	lá	à-gbé	shìn	ngèr
Child	def	it-fall	into	water

The child fell into the water.

Negation, even in complex sentences, is marked at the very end of a sentence:

m̀.fá	àlù	wán	lá	à-kpé	gá.
I-know	if	child	DEF	it-die	NEG

I don't know whether the child died.

In some cases, as in the following focus construction, negation is additionally marked on the verb:

tùɔ TM	ù	wán	lá	(à)-ne-ná	àyàbà	gá
Chief	REL	child	DEF	(it)-NEG-give	plantain	NEG

It's not a chief the child gave the plantain.

Relativization is marked by a REL ative pronoun indicating gender and number at the beginning and a low tone at the end of the relative sentence. The low tone has the effect of down-stepping the following high-toned syllable. If the relativized noun is marked with the invariable definitizer **lá**, the relative clause appears between the head noun and the **lá**. If, moreover, the relativized noun is definite, a resumptive pronoun appears in the 'original' syntactic position in the relative clause:

Wán	ù	m̀-méngé	`	ndí
Child	REL	i-see	REL	lost

A child I saw is lost.

Wán	ù	m̀-méngé	àná	lá	`	ndí
Child	REL	i-see	PRO	DEF	REL	lost

The child I saw is lost.

In relativized objects of prepositions, a shortened form of the resumptive pronoun appears as well. The preposition maintains its original syntactic position:

(à)-lú	shìn	yò	ì	wán	lá	gbé	shìn	mì	`.
(he) live	in	house	REL	child	DEF	fall	into	pro	REL

He lives in the house the child fell into.

There is no clear evidence of serialization. Complex structures giving the appearance of serialization are instances of simple conjoined sentences:

wán	lá	à-té	àyàbà,	à-dzé	từɔ̀ TM	ámì.
child	DEF	it-take	Plantain	it-go	to chief	with them

The child took the plantain (and) went to the chief with them.

Only in cases where verbs of motion that do not require a goal are involved does one obligatorily find **dze** 'go' as the second verb. The latter then requires the expected preposition:

wán	lá	à-yévísí	dzé	shìn	yò
child	DEF	it-run	go	into	house

The child ran into the house.

Comparatives have the Africa-typical structure: X exceeds Y in Z, or X is Z, exceeds Y:

unòr(-mban)	bàìkɔ́	hèmbà	ùnor-ngèr
elephant	fat	exceed	hippopotamus

An elephant is fatter than a hippo.

unòr(-mban)	hèmbá	bàìkɔ́-n	ùnor-ngèr
elephant	exceeds	being fat	hippopotamus

An elephant is fatter than a hippo.

There are a good number of ideophones; not all of them, however, are semantically 'identical' to the head verb:

gbúká	lá	gbà	pùú
gun	DEF	fire	BANG

The gun went off 'bang'.

In this case pùú is the sound made by the gun and only that. In the following two sentences, however, **pùlùtùtù** and **cìcá** can be used with any color to express intensity:

àvámbè	nìà	pùlùtùtù cìcá
blood	red	INTENSE

blood is absolutely red

àlísì	yìlè	pùlùtùtù cìcá
charcoal	black	INTENSE

charcoal is absolutely black

Tivoid. Aside from a very large number of vocabulary resemblances, all languages belonging to the Tivoid subgroup share a complex nominal system where a noun is marked both by a prefix and a suffix. A single example—the word for *oil*—suffices to show the unity of Tivoid on the one hand and the close, but non-Tivoid character of neighboring Bantoid languages on the other:

Tivoid		other Bantoid	
BATOMO	mákutúm	AMANAVIL	mètí
BELEGETE	mutum	AMASI	òmìlí
ICHEVE	mukurem	AMBELE	mùút
NDII	mkutem	ASAKA	ɔ́múú
OLITI	mùkùlémí	ASUMBO	mɔɔ́t
TIV	ḿkúrém	ESIMBI	mìri
TIVI	íŋkùlŋ	KENYANG	bawet
YIIVE	mǝkùtèm	MESAKA	òmbúl

References

Abraham, Roy Clive. 1933. The grammar of Tiv, Vol. 5(1), 213pp. Kaduna: Government Printer.

———. 1940a. The principles of Tiv, 102pp. London: Crown Agents.

———. 1940b. Dictionary of the Tiv language, ix, 331 pp.London: Crown Agents. [reprinted 1968 by Gregg].

———. 1940c. A Tiv reader for European students, 82pp. London: Crown Agents. [Appendix: Further points of grammar for reference, pp. 69–82].

Arnott, David W. 1958. The classification of verbs in Tiv. Bulletin of the School of Oriental and African Studies 21(1). 111–33.

———. 1964. Downstep in the Tiv verbal system. African Language Studies 5.34–51.

———. 1967. Some reflections on the content of individual classes in Fula and Tiv. La Classification Nominale dans les Langues Négro-Africaines, ed. by Gabriel Manessy, 45–74.

———. 1992. R. C. Abraham's books on Tiv. In Jaggar, pp. 141–5.

Bohannan, Laura, and Bohannan, Paul. 1953. The Tiv of Central Nigeria, viii, 9-100. London: International African Institute (Ethnographic Survey of Africa, Western Africa, Part 8).

Clarke, John. 1848. Specimens of dialects: short vocabularies of languages: and notes of countries & customs in Africa, 104pp. Berwick upon Tweed: D. Cameron. [Edited with a commentary by Edwin Ardener, with a biographical note by Shirley Ardener. Gregg Press, 1972, 76pp.].

Jagger, Philip J. (ed.) 1992. Paper in Honour of R.C. Abraham (1890–1963). London: School of Oriental and African Studies.

Jockers, Heinz. 1991a. Studien zur Sprache der Tiv in Nigeria, xv, 164pp. Frankfurt am Main: Peter Lang. (Europäische Hochschulschriften 21).

———. 1991b. The alphabetization of the Tiv: a case study. Language standardization in Africa/Sprachstandardisierung in Afrika/Standardisation des langues en Afrique, ed. by Norbert Cyffer et al., 145–56. Hamburg: Helmut Buske.

———. 1992. Linguistic studies on Tiv—before, by, and after R. C. Abraham. In Jaggar, pp. 163–73.

Koelle, Sigismund Wilhelm. 1854. Polyglotta africana; or a comparative vocabulary of nearly three hundred words and phrases in more than one hundred distinct African languages, vi, 24, 188 p. London: Church Missionary House. [Unveränderter Nachdruck der Ausgabe London 1854 vermehrt durch eine historische Einführung von P. E. H. Hair und einen Wortindex von David Dalby. Graz, Akademische Druck- u. Verlagsanstalt, 1963. 19pp.].

Lukas, Johannes. 1952. Das Nomen im Tiv. Anthropos 47.147–76.

Malherbe, W. A. 1932. Tiv–English dictionary, with grammar notes and index, xxxix, 207 pp. Lagos: Government Printer.

Sibomana, Leonidas. 1980. Aspects of Tiv morphophonemics. Afrika und Übersee 63(1).69–77.

———. 1981. Tonal lexicalization of Tiv nouns. Festschrift zum 60, ed. by Inge Hofmann, Hrsg. Geburtstag von P. Anton Vorbichler, 1. Teil, 155–73. Wien: Institut für Afrikanistik und Ägyptologie der Universität Wien.

———. 1983. Tonal structure of Tiv verbs. Afrika und Übersee 66(1).77–89.

———. 1984. Displacement and lowering of Tiv tones. Afrika und Übersee 67(2).277–87.

Terpstra, Gerard. 1968. English–Tiv dictionary, 120pp. Ibadan: University of Ibadan. (Institute of African Studies, Occasional Publications No. 13).

F. K. ERHARD VOELTZ

Tocharian

Tocharian is the name applied to a small group of extinct Indo-European language varieties attested in fragmentary manuscripts first discovered in the last decade of the nineteenth century CE. The recovery of manuscripts in Tocharian as well as others written in Khotanese Saka, Sanskrit, and various Turkic and Sino-Tibetan languages from the Taklamakan Desert, Tarim Basin, and region south of the Tian Shan range, all located in the modern-day Xinjiang Uygur Autonomous Region of northwestern China, was a consequence of the many exploratory expeditions undertaken around the turn of the century along the ancient Silk Road trade routes. The majority of extant Tocharian texts originate, by most estimates, from the sixth to ninth centuries CE, after which the Tocharian-speaking culture is believed to have assimilated with neighboring (e.g. Turkic) cultures or died out.

The texts, written on palm leaves, paper scrolls, and wooden tablets, represent a diverse array of genres and subject matter, from scientific, legal, and commercial documents (such as caravan passes) to dramas, narratives, poetry, and Buddhist and Manichaean religious works. A modified form of the Brāhmī script of northern India, the writing system employed in the Tocharian texts, consists of a syllabary in which each consonantal letter is paired with an inherent (unwritten) vowel. Vowels other than this inherent vowel are indicated with distinctive letters or diacritics.

The first extensive analysis and decipherment of the hitherto unknown Tocharian language were made by Emil Sieg and Wilhelm Siegling in 1908. The decipherment was facilitated by the knowledge of other languages written in Brāhmī and related syllabaries as well as the existence of bilingual Buddhist religious texts (in the original Sanskrit, accompanied by Tocharian translations). It is still under debate whether the name Tocharian itself is the most proper name for this language (group); this topic alone has yielded much speculation about the origins of Tocharian and its speakers. The name Tocharian is associated with a people called the Tocharoi, who, as suggested in the writings of the ancient Greeks Strabo and Ptolemy, at one time inhabited the area formerly known as Bactria in present-day Afghanistan and Tajikistan. Although the existence of the Tocharoi is not in question, some scholars believe that they were of Iranian stock and bore no relation to the speakers of the non-Iranian language in the Tocharian manuscripts. Still others who may not necessarily share this belief maintain that the name *Tuγri* (*Twγry/Twghry/Twqri*) found in the neighboring Turkish language Uygur would be more appropriate as a name for Tocharian. Despite the conflicting views among scholars on these matters, however, the name Tocharian has prevailed within the general community of linguists.

Apart from their association with the Tocharoi, it has also been suggested that the speakers of Tocharian were one and the same with a culture that the Chinese called the Yuezhi, a confederacy of several (possibly Indo-European-speaking) tribes originally living at the edge of the Gobi Desert in present-day Gansu Province of northern China and not overly distant from the sites of the Tocharian discoveries. Having suffered a major military defeat in the middle of the second century BCE, evidence from Chinese and Greek annals as well as from sites along the migration route exhibiting Yuezhi-particular burial practices corroborates the view that Yuezhi tribes journeyed west out of China, deeper into central Asia. By the early 130s BCE, some of the Yuezhi tribes had arrived in Bactria and neighboring Sogdiana surrounding present-day Samarkand in Uzbekistan. These tribes that had reached Bactria are known as the Greater Yuezhi. At around 50 CE they may have established the Kushan Empire that extended into northern India. Some generally accepted views are that neither the Tocharoi of Greek sources nor the Greater Yuezhi were the speakers behind the Tocharian manuscripts. Others suggest that both peoples were actually one and the same since both ended up in Bactria and appeared to have had otherwise similar histories according to the written records available. These are fair assumptions given the fact that by the early second century BCE the Greater Yuezhi had already migrated well past the sites of the Tocharian manuscripts first produced approximately 650–750 years later.

At least one more possibility for a Yuezhi-Tocharian connection exists, however. The Lesser Yuezhi, who evidently halted their migration out of more easterly districts of China well short of the trek completed by their Greater Yuezhi kinsmen to Bactria, may have settled south of the Tarim Basin and Taklamakan Desert for a time. Between the cities of Niya and Loulan in this area (both buried by sand over the ages), documents in the Indic language Niya Prakrit came to light and have been claimed to demonstrate signs of influence from a Tocharian language, which may have existed in the vicinity at least several centuries prior to the period of the Tocharian manuscripts found further north. It is possible that the (Lesser) Yuezhi, if not of Iranian origin, as commonly believed, were the speakers of this hypothetical southern Tocharian language and eventually migrated north where they gave rise to the speakers of the attested manuscripts. If correct, this theory would serve to link the Greater Yuezhi, as kinsmen of the Lesser Yuezhi, with one or more Tocharian languages. New manuscript discoveries along the migratory paths of the Greater Yuezhi to Bactria would clarify this view.

It has been suggested that Tocharian was employed as a lingua franca along the Silk Road and other trade routes in the region among people who spoke mutually unintelligible languages. However, most of the Tocharian manuscripts—in two distinct Tocharian varieties no less (A and B)—have been concentrated within a single region and little to no evidence of a significant Tocharian presence beyond the confines of the Tarim Basin has appeared to date to confirm this view. Whatever their origins, the speakers of Tocharian appear to have filled many cultural and economic roles in northwestern China for at least three centuries from the mid to late first millennium CE. They played a considerable role, for example, in the spread of Buddhism and Manichaeism (with roots in northern India and Persia, respectively) into western China. Trade and other forms of contact with the Chinese, Persians, Turks, various cultures from India, and others secured for the Tocharian speakers of the Tarim Basin sites a position of great influence, which endured for some time.

From a linguistic standpoint, scholars were at once amazed and perplexed after it was determined that Tocharian was in fact Indo-European. A traditional and previously straightforward way of grouping the Indo-European languages was to refer to the *centum/satem* divide. Languages of the *centum* group (so-called after the Latin word for '100') such as Celtic, Germanic, Greek, and Italic have velar reflexes (or later correspondents) of velar sounds (e.g. /k/ and /x/) of the reconstructed Proto-Indo-European (PIE) language, while languages of the *satem* group (so-called after the Avestan word for '100') such as Armenian, Indo-Iranian, and Slavic have sibilant reflexes (e.g. /s/ and /š/). Since the word for '100' is *känt* in Tocharian A and *kante* in Tocharian B, the once solid west/east grouping seen in the *centum/satem* divide was disturbed.

With regard to geographical and social distinctions among the varieties of Tocharian, which some consider separate languages, others merely dialects of a single language, Tocharian A was found concentrated in the oasis town of Turfan and to a lesser extent further west at the central site of Karashahr, while Tocharian B was mainly concentrated in Kucha, the westernmost of the three primary sites, and found to some extent in both Karashahr and Turfan as well. Based on these geographical differences, Tocharian A has also been called East Tocharian, Turfanian, Karashahrian, and Agnean (after the former state of Agni situated around Karashahr), while Tocharian B has also been referred to as West Tocharian and Kuchean (after the former state of Kuchi and its site around Kucha). Under one proposal, both varieties of Tocharian were living languages, with B having been imported to the two eastern sites by Buddhist missionaries. A second suggests that A was a dead language used primarily for liturgical purposes

while B was the living language used in everyday life since it was common to all three sites.

While the status of Tocharian as a member of Indo-European remains uncontroversial, the exact relationship between the Tocharian languages and other Indo-European branches has long been subject to speculation. Linguists have made attempts to establish genetic ties between Tocharian and each of the Anatolian, Balto-Slavic, Celtic, Germanic, Greek, Indo-Iranian, and Italic branches at one time or another—in other words, between Tocharian and most of the Indo-European subgroupings. This only serves to demonstrate the relatively modest consensus that exists over the complex position occupied by Tocharian within the Indo-European context in historical, linguistic, and social terms.

The phonology (sound structure) of Tocharian is characterized by palatalized variants of almost all consonants, a single series of stop consonants (i.e. voiceless /p/, /t/, /k/ but no voiced /b/, /d/, /g/ or voiced aspirated correspondents—all three series existed in PIE), the loss of final consonants in A and B (e.g. PIE *tod → B te 'this'/'it') and final vowels in A (e.g. klots 'ear' vs. B klautso, lānts 'queen' vs. B lāntsa, wäl 'king' vs. B walo), and some loss of PIE distinctions between short and long vowels. Singular and plural, as well as dual (e.g. B wi rsoñc 'two spans' vs. the plural form rsonta) and further subtleties in number distinction, are also characteristic of Tocharian. A unique feature of Tocharian among the Indo-European languages is its elaborate system of several primary cases (e.g. nominative and genitive) and a host of secondary cases such as the perlative, allative, and comitative, each of which has a specific grammatical function. The secondary cases are formed by the suffixation of various postpositional elements to forms in the oblique, one of the primary cases. Within the verbal system, both active and medio-passive voices are found. The latter is marked in the present tense by an -r ending (e.g. A klyosamtär and B klyausemtär 'we hear') as in some other centum languages and Indo-Iranian from the satem group. Lexical borrowings from Sanskrit and neighboring Iranian languages are common. Tocharian is thought to have been influenced lexically and structurally, to various extents, by the myriad of non-Indo-European languages in northwestern China and its periphery, such as Sino-Tibetan and Turkic, as well as those encountered by the ancestors of the Tocharian speakers along the eastward migratory paths out of the central Indo-European homeland and eventually into China (e.g. Finno-Ugric languages).

The harsh, arid environment around the Taklamakan Desert region, which was ideal for safeguarding the Tocharian manuscripts for up to 1,500 years, was also instrumental in preserving several groups of mummified humans first discovered during the Western expeditions of the late nineteenth and early twentieth centuries CE. These mummies, however, have only been widely known and studied in the last quarter of the twentieth century CE. Due to their blonde and reddish hair and other distinctive bodily features, some scholars claim that these mummies were the ancestors of the Tocharian speakers. Clearly, they are not easily identifiable with most of the cultures from the past or present that are commonly linked to the region, such as the Chinese, Mongols, and Turks, but rather have appearances more typical of Europeans. Additional clues such as wall paintings found in caves around Kucha and Turfan depicting humans with red hair, along with the proximity of the mummy sites to the sites of the Tocharian manuscripts, all serve to fuel the proposed connection of the mummies with the speakers of Tocharian.

The discovery of Tocharian has influenced many Indo-Europeanists to reconsider their views about the history of the Indo-Europeans and the development and interrelations of the languages attributed to them. Various forms of linguistic and archeological evidence may serve to further unlock the secrets presented by the speakers of Tocharian A and B (and perhaps other varieties of Tocharian that have left little to no trace).

Among the major figures whose works are not listed in the References section but who have nonetheless made substantial contributions to the advancement of knowledge about the origins of Tocharian and its speakers are Elizabeth J.W. Barber, Gerd Carling, Walter Couvreur, Viacheslav V. Ivanov, Jay H. Jasanoff, Sylvain Lévi, Victor H. Mair, James P. Mallory, Antoine Meillet, Holger Pedersen, John H.W. Penney, Pavel Poucha, Ernst Schwentner, and Werner Winter.

References

Adams, Douglas Q. 1988. Tocharian historical phonology and morphology. New Haven, CT: American Oriental Society.

Adams, Douglas Q. 1999. A dictionary of Tocharian B. Amsterdam, Atlanta: Rodopi.

Krause, Wolfgang. 1952. Westtocharische Grammatik (=West Tocharian Grammar). Heidelberg: Carl Winter.

Krause, Wolfgang, and Werner Thomas. 1960–1964. Tocharisches Elementarbuch (=Elementary Tocharian book), 2 vols. Heidelberg: Carl Winter.

Lane, George S. 1966. On the interrelationship of the Tocharian dialects. Ancient Indo-European dialects, ed. by Henrik Birnbaum and Jaan Puhvel. Berkeley, CA: University of California Press.

Lane, George S. 1967. On the significance of Tocharian for Indo-European linguistics. Studies in historical linguistics in honor of George Sherman Lane, ed. by Walter W. Arndt, et al. Chapel Hill, NC: University of North Carolina Press.

Lane, George S. 1970. Tocharian: Indo-European and non-Indo-European relationships. Indo-European and Indo-Europeans,

ed. by George Cardona, Henry M. Hoenigswald, and Alfred Senn. Philadelphia: University of Pennsylvania Press.

Pinault, Georges-Jean. 1989. Introduction au tokharien (=Introduction to Tocharian). Lalies: actes des sessions de linguistique et de littérature de l'École normale supérieure 7. 5–224.

Ringe, Donald A. 1991. Evidence for the position of Tocharian in the Indo-European family? Die Sprache 34. 59–123.

Ringe, Donald A. 1996. On the chronology of sound changes in Tocharian, Vol. I. New Haven, CT: American Oriental Society.

Sieg, Emil, Wilhelm Siegling, and Wilhelm Schulze. 1931. Tocharische Grammatik (=Tocharian Grammar). Göttingen: Vandenhoeck & Ruprecht.

Windekens, A.J. van. 1976–1982. Le tokharien confronté avec les autres langues indo-européennes (=Tocharian confronted with the other Indo-European languages), 2 vols. in 3. Louvain: Centre international de dialectologie générale.

SEAN O'ROURKE

See also **Indo-European 1: Overview**

Tok Pisin

Tok Pisin (or New Guinea Pidgin) is the dialect of Melanesian Pidgin spoken in Papua New Guinea. It serves as the main language of wider communication in a country where more than 800 separate indigenous languages are spoken by a population of nearly five million. The two other dialects of Melanesian Pidgin are Pijin, spoken in the Solomon Islands (with more than 80 indigenous languages and a population of around 480,000), and Bislama, spoken in Vanuatu (more than 100 languages, population 192,000). Torres Strait Creole (also known as Broken or Yumiplatok)—spoken by approximately 10,000 people around the northern tip of eastern Australia—is closely related to Melanesian Pidgin but is usually considered to be a separate language.

Melanesian Pidgin is an example of a contact language—a new language that developed in a situation in which speakers of different languages needed to communicate but did not share a common language. Most of the vocabulary comes from one language, called the 'lexifier' (English, in the case of Melanesian Pidgin), but the words often have different meanings. For example, in Tok Pisin, *spak,* derived from English *spark,* means 'drunk'. Many words are also derived from other languages (see below). Furthermore, the phonology and morphosyntax of a contact language such as Tok Pisin are distinct from those of the lexifier. For example, in phonology, Tok Pisin has a five-vowel system (a, e, i, o, u) and only 15 consonants (represented in writing as b, d, g, h, k, l, m, n, p, r, s, t, v, w, and y). The English phoneme inventory contains about twice as many elements. For sentence structure and meaning, Tok Pisin uses separate words to show number and tense (see examples below) where English generally uses suffixes, e.g. *-s* and *-ed.* However, Tok Pisin has a few suffixes not found in English, such as *-pela* on adjectives, e.g. *bikpela haus* 'big house' and *-im* on transitive verbs (see below).

Such a contact language is called a 'pidgin' when it continues to be used primarily as a second language for intergroup communication, and it is called a 'creole' when it becomes the mother tongue of a particular community of speakers. As will be seen below, there is some controversy about whether Tok Pisin is actually a pidgin or a creole.

Because Melanesian Pidgin is a contact language that came into being fairly recently, we can describe almost its entire history. The first stage of the development of Melanesian Pidgin dates from the early 1800s, when Melanesians began to have frequent contact with Europeans (including Australians and Americans). This was the result of whaling in the region, followed by trading in sandalwood and bêche-de-mer. Because Melanesia is one of the most linguistically diverse areas of the world, it was impossible for Europeans to learn the local languages for trading (as they did in other areas of the Pacific). Thus, to communicate, they used simplified English or existing contact languages, such as South Seas Jargon and various forms of Aboriginal Pidgin English from Australia. As a result of these limited encounters, many Melanesians picked up some English vocabulary and phrases from the existing contact languages.

The second stage came with the beginning of the Pacific labor trade in 1863, when Melanesians started to be recruited (and in some cases kidnapped) to become laborers for plantations in Queensland (Australia) and Samoa. Melanesians from diverse areas found themselves literally in the same boat, and their only common language was what they had acquired from earlier contacts with Europeans. Thus, they used this to communicate with one another on the ships and later on the plantations. With continued use, new features were

added, norms began to emerge, and a stable pidgin language began to develop—early Melanesian Pidgin.

The earliest Queensland laborers were mainly from the New Hebrides (now Vanuatu) and the Solomon Islands. Some laborers from German New Guinea also went to Queensland in 1883 and 1884, but many more went to plantations in German-controlled Samoa, from 1879 to 1912. Laborers from the other countries also started going to Samoa in 1878, and many of these had already worked in Queensland. Thus, early Melanesian Pidgin was transported to Samoa. However, after 1885, laborers from the New Hebrides and Solomon Islands were no longer recruited for Samoa, and early Melanesian Pidgin began to diverge into two slightly different varieties—one spoken in Queensland and one in Samoa.

The third stage of development began when laborers returned to their home islands, after their contracts finished, and brought the developing pidgin with them. Previously, these islands had no language of wider communication; because the pidgin served this function well, it spread rapidly. It was also used by the large-scale internal labor force that worked on the plantations of German New Guinea, the New Hebrides, and the Solomon Islands after the external labor trade had ended at the turn of the century. In each of these countries, early Melanesian Pidgin further stabilized and changed under the influence of the local indigenous languages. Today, the three dialects differ mainly in vocabulary and a few grammatical rules (see below).

After Tok Pisin had stabilized, it began to be used for new functions, such as religion. It was developed into a written language by missionaries in the 1930s and was later used in newspapers and radio broadcasting. As its use was extended into new areas, it changed linguistically to become more complex; i.e. it acquired more vocabulary and more grammatical rules and inflections. (The same process occurred with Bislama and Pijin.) Thus, in both function and structure, Tok Pisin (and Melanesian Pidgin as a whole) became what is called an 'expanded pidgin'.

Finally, in recent years, especially in urban areas of Papua New Guinea, such as Port Moresby and Lae, many people have been marrying outside their traditional language groups. Thus, often the common language of the parents is Tok Pisin, and this is what their children acquire as their first language. Because of this nativization (the process of a pidgin becoming a native language), some linguists apply the term 'creole' to Tok Pisin (and Melanesian Pidgin in general), emphasizing that it has thousands of native speakers and has the functions and grammatical features found in typical creoles. Those who say that it is still a pidgin demonstrate that more than 90% of its speakers have learned it as a second rather than a first language and

that nearly all of the Tok Pisin-speaking population are bi- or multilingual, whereas creole-speaking populations are most often monolingual.

Today Tok Pisin is the lingua franca of the entire country of Papua New Guinea; it is known by an estimated three quarters of the country's nearly five million inhabitants. It is the most widely used language of urban areas and is also widely spoken in rural areas in the northern part of the country.

Tok Pisin is used to some extent in radio and television broadcasting, especially in interviews and news reports. (It is also used in some of Radio Australia's Papua New Guinea Service broadcasts.) The weekly Tok Pisin newspaper *Wantok* has a readership of more than 10,000, and many government publications are also written in Tok Pisin. The language is widely used in religion, and there is a Tok Pisin translation of the New Testament of the Bible.

The constitution of Papua New Guinea recognizes Tok Pisin as one of the national languages. (The constitution of Vanuatu names Bislama as the national language.) Although English is more widely used for government business, much of the debate in Parliament is in Tok Pisin.

Until recently, English was the official language of education in Papua New Guinea, and it was used in all government schools (although Tok Pisin was widely used in community and church-run preschools and vocational schools). However, with the recent education reform, communities can choose the language to be used in the first three years of elementary education, and many have chosen Tok Pisin. Although many people still feel that Tok Pisin is inferior to English, most accept it as a separate language and an important language of Papua New Guinean identity.

Tok Pisin differs from the other two dialects because, as mentioned above, most of the returned laborers had worked in Samoa rather than Queensland. In addition, nearly all of the New Guinea laborers were from New Britain and New Ireland and the neighboring small islands, where the internal German-owned plantations were also located. Thus, Tok Pisin has many words from the languages of these islands (especially Tolai), as well as some from Samoan and German. Some examples are *kiau* 'egg' and *lapun* 'old' from Tolai, *malolo* 'rest' from Samoan, and *gumi* 'rubber', *beten* 'pray', and *raus* 'get out!' from German.

Bislama has many words derived from the languages of central Vanuatu, e.g. *nakamal* 'meeting house' and *nabanga* 'banyan tree'. Bislama also differs from both Tok Pisin and Pijin in having many words derived from French. (Both France and Britain were the former colonial powers in Vanuatu.) Some examples are *bonane* 'New Year's celebration', *pima*

'chilli', and *lafet* 'holiday'. Pijin differs from the other two dialects in having very few words derived from local languages or any language other than English.

An example of a grammatical difference between the three dialects is in the way they indicate progressive aspect (marked with *-ing* in English). Tok Pisin most often uses *wok long* (literally 'work at') before the verb or the locative or existential verb *stap* 'stay, exist' after the verb (with an intervening *i*); Bislama also uses *stap* but before the verb; and Pijin sometimes uses initial consonant–vowel reduplication (but often does not mark progressive aspect). For example, 'I'm drinking' would be the following in the three dialects: *Mi wok long dring* or *Mi dring i stap* (Tok Pisin); *Mi stap dring* (Bislama); and *Mi didring* (Pijin).

Tok Pisin has been the focus of research testing several theoretical questions in creolistics (the study of pidgin and creole languages). The first concerns 'typical' creole features and developmental history. Creole genesis (or 'creolization') in areas such as the Caribbean is believed to have occurred very quickly, with stabilization and expansion occurring along with nativization. However, the creolization of Tok Pisin has been taking place gradually, with stabilization and then expansion occurring before nativization. Nevertheless, it shares many features with other creoles, such as the following:

(1) Tense, aspect, and modality are indicated by separate words occurring before the verb, e.g. *Em i bin wok asde,* 'he worked yesterday' and *Em bai i wok tumora,* 'he will work tomorrow'.
(2) Lack of a copula (a form of the verb 'to be'), e.g. *mi hanggre,* 'I'm hungry'.
(3) Use of the third-person plural pronoun (*ol* in Tok Pisin) as a plural marker, e.g. *mi lukim ol dok,* 'I saw the dogs'.

Second is the theoretical question of discontinuity. It has been claimed that creoles are structurally discontinuous from their pidgin predecessors in having many more complex structures. These are supposedly innovations of children that occurred during nativization and could be attributed to their innate language faculty. Tok Pisin differs from nearly all known creoles in that it is still acquiring native speakers, and therefore it has provided a good opportunity for studying creolization as it occurs. Because it was believed that grammatical expansion in creoles is the result of nativization, researchers expected to find discontinuity between the Tok Pisin spoken as a second language by adults and the creolized Tok Pisin of children who acquired it as a first language. However, no substantive differences have been found.

Last is the question of substrate influence—the degree to which the first languages of the originators of the pidgin or creole (the substrate or substratum languages) have affected its structure. Although many scholars have downplayed the role of substrate influence in pidgins and creoles, many features of Melanesian Pidgin are clearly derived from or reinforced by those of the Oceanic Austronesian substrate languages. These include the following:

(1) The transitive suffix (showing that the verb takes a direct object) (*-im* in Tok Pisin), e.g. *meri i bin boilim wara,* 'the woman boiled the water'.
(2) The subject referencing marker *(i)* in the verb phrase, e.g. *Tom i wok long Mosbi,* 'Tom works in Port Moresby'.
(3) Exclusiveness and dual and trial number marked in the pronoun system, e.g. *yumi* (first-person plural inclusive, i.e. 'we, including you') vs. *mipela* (first-person plural exclusive, i.e. 'we, not including you') and *yu* (second-person singular, i.e. 'you, one person'), *yutupela* (second-person dual, i.e. 'you, two persons'), *yutripela* (second-person trial, i.e. 'you, three persons'), and *yupela* (second-person plural, i.e. 'you, many people').

References

Crowley, Terry. 1990. From beach-la-mar to Bislama: the emergence of a national language in Vanuatu. Oxford: Clarendon and New York: Oxford University Press.

———. 2003. A new Bislama dictionary, 2nd edition. Suva, Fiji: Institute of Pacific Studies, and Vila, Vanuatu: Pacific Languages Unit (University of the South Pacific).

Heubner, Thom, and Stephen Rex Horoi. 1979. Solomon Islands Pijin. Brattleboro, Vermont: The Experiment in International Living.

Jourdan, Christine. 2002. Pijin: a trilingual dictionary. Canberra: Australian National University (Pacific Linguistics).

Keesing, Roger M. 1988. Melanesian Pidgin and the Oceanic substrate. Stanford: Stanford University Press.

Mihalic, F. 1971. The Jacaranda dictionary and grammar of Melanesian Pidgin. Milton, Queensland: Jacaranda.

Romaine, Suzanne. 1992. Language, education, and development: urban and rural Tok Pisin in Papua New Guinea. Oxford: Clarendon.

Shnukal, Anna. 1988. Broken: an introduction to the Creole language of Torres Strait. Canberra: Australian National University (Pacific Linguistics C-107). reprint, 1998.

Siegel, Jeff. 1998. Substrate reinforcement and dialectal differences in Melanesian Pidgin. Journal of Sociolinguistics 2.

Simons, Linda, and Hugh Young. 1978. Pijin blong Iumi: a guide to Solomon Islands Pijin. Honiara, Solomon Islands: Solomon Islands Christian Association.

Smith, Geoff P. 2002. Growing up with Tok Pisin: contact, creolization, and change in Papua New Guinea's national language. London: Battlebridge.

Wurm, S.A., and P. Mühlhäusler (eds.) 1985. Handbook of Tok Pisin (New Guinea Pidgin). Canberra: Australian National University (Pacific Linguistics C-70).

JEFF SIEGEL

Tone Languages

In addition to consonants and vowels, languages also use the suprasegmental to convey different meanings. Through stress, English contrasts such pairs of words as /ˈinsult/ vs. /inˈsult/ and /ˈpermit/ vs. /perˈmit/. Many languages employ tone for similar purposes, but they differ from stress languages in some significant ways. First of all, tone languages regard tone as an important means to contrast lexical meanings. Thus, Mandarin distinguishes unrelated words by tone alone (the raised numerals indicate relative pitch levels at which a tone starts and ends, with ^{5} representing the highest and ^{1} the lowest): /aj^{55}/ 'sadness' vs. /aj^{35}/ 'suffer' vs. /aj^{214}/ '(of height) short' vs. /aj^{51}/ 'love'. (cf. the English examples above, where stress contrasts only word categories—nouns vs. verbs). Second, the occurrence of tone in a word is not predictable, whereas the assignment of stress is rule-governed. Therefore, we may define tone language as one that uses distinctive pitch to contrast concrete meanings of simple words, with the assignment of tone unpredictable by phonological rules.

Lexical tone and stress need not be regarded as a pair of dichotomies, however. Indeed, some languages make use of both. While no European language is predominantly tonal, most dialects of Norwegian, Swedish, Slovene, Serbo-Croatian, and Lithuanian have prosodic systems with a combination of tone and stress. In these Scandinavian languages, tone assignment is sensitive to stress, whereas the reverse applies to the South Slavic and Baltic languages—the stress assignment is said to be sensitive to tone.

Based on distinctive organizations of tones, three major types of tone language can be recognized, i.e. syllable-tone, word-tone, and pitch-accent languages. Syllable-tone languages represent the best-known type of tone system, in which the syllable constitutes the basic domain for each tonal category. Cantonese is an excellent example of a syllable-tone language, e.g. /fu^{55}/ 'call aloud', /fu^{35}/ 'tiger', /fu^{33}/ 'rich', /fu^{21}/ 'hold', /fu^{13}/ 'woman', and /fu^{22}/ 'father'. The tone may change, under certain circumstances, from one category to another. However, individual tonal components of a category do not spread freely in syllable-tone languages. Only rarely do we see the copying of an entire tonal category from one syllable onto an adjacent one in such languages.

Taking the simple word as its basic domain, word-tone languages assign tonal categories to the word as a whole, irrespective of the number of syllables available. Core tonal categories can exhaustively be identified in shorter words. Mende (an African language) has a word-tone system. The actual patterns for words carrying the tonal category of LH (a low tone plus a high tone, to be signified with the diacritics ˋ and ´, respectively) thus vary as follows: /mbǎ/ 'rice', /fàndé/ 'cotton', /ndàvúlá/ 'sling'. The LH pattern is observed only in disyllabic words; it becomes a rising tone in monosyllabic words, whereas in trisyllabic words the H tone spreads onto the final syllable. Tone-spreading within the basic domain of tonal categories is common.

Pitch-accent languages such as Japanese generate tonal patterns in an abstract domain, which may be called 'templet'. They do not directly contrast one tone against another. Instead, tonal categories emerge from the orderly placement of a marked tone at the nth position in the templet, where n typically starts from 0 or 1 and continues up to n. Tone-spreading may occur within the templet to increase the repertoire of tonal categories. As the pattern of a tonal category is predetermined in the templet, situational neutralization of two tonal categories is possible in shorter words. For instance, the citation form of Tokyo Japanese /hàná/ is ambiguous between 'nose' and 'flower', while in fact the two carry different tonal categories. The neutralization ceases when the nouns are attached with the topic clitic /wa/: /hàná wá/ 'as for noses' vs. /hàná wà/ 'as for flowers'. The major characteristics of these three types of tonal system are summarized in Table 1.

TABLE 1 A Comparison of Syllable-Tone, Word-Tone, and Pitch-Accent Languages

	Syllable-Tone	Word-Tone	Pitch-Accent
Basic domain of tonal categories	Syllable	Word	Templet
Manifest of all tonal categories in shorter words	Yes	Yes	Not necessary
Neutralization of tonal contrasts in the citation form	Impossible	Not possible	Possible
Spreading of individual tone component	Rare	Common	Common

Tone Languages of Asia

Tone languages are abundant in Asia, stretching from Japan in the east to Nepal in the west, and from China in the north to Thailand and Myanmar (Burma) in the south. Sporadic tone languages are also found in Indonesia and Pakistan. Sino-Tibetan comprises the largest family of tone languages in this region, but some Tibeto-Burman languages have no tones, e.g. Amdo Tibetan and the Kiranti languages of Nepal. All of the Tai-Kadai (including Thai and Lao) and Hmong-Mien (Miao-Yao) languages, spoken in Southeast Asia and southern China, are tonal. Nontonal Southeast Asian languages are mainly those of the Austronesian (including Malay, Indonesian, and Pilipino) and Austroasiatic families, with the notable exception of Vietnamese.

Tone is well developed in all Chinese languages. Classic works describe Middle Chinese as having four tonal categories: Level, Rising, Departing, and Entering. Later linguistic development has brought tonal systems of Chinese to extreme diversity. Among Mandarin dialects, for example, some have more tones than Beijing, while others have fewer. Furthermore, pitch values for the same tonal category may also vary from one dialect to another, as shown in Table 2.

Korean represents an unusual case of dialectal variation in tone. While the Central dialect of Seoul is no longer tonal, tone distinctions in Middle Korean have been retained in Northeastern (Hamkyeng) and Southeastern (Kyengsang) dialects. The latter has such contrast as /mál/ 'horse' vs. /māl/ 'measure' vs. /màl/ 'word'.

The majority of tone languages in China and Southeast Asia are of the type of syllable-tone. Word-tone languages are found mainly in Tibeto-Burman languages of Tibet and Nepal, including Ü (Lhasa Tibetan), Tamang, Thakali, and Gurung. In addition, Shanghainese has also been identified as a type of word-tone language.

The best-known pitch-accent language is Tokyo Japanese. Apart from toneless dialects such as Sendai, all Japanese dialects have pitch-accent systems—some simpler while others more elaborate. Another pitch-accent language is Prinmi (Pumi), a Tibeto-Burman language of southwestern China. A three-way contrast exists in monosyllabic words, e.g. /ʃî/ 'louse' (high-falling) vs. /ʃí/ 'hundred' (high-level) vs. /ʃǐ/ 'new' (rising), but the first two words are nearly neutralized in the citation form. Tonal distinctions in Prinmi are made by specifying the locus of the H tone in the templet and whether it spreads to the next unit. Expressing these two parameters as [n] [±spread], the full patterns of the seven tonal categories of Prinmi are illustrated as in Table 3.

Tone Languages of Africa

The great majority of African languages south of the Sahara have tone. Except for Fulani, Wolof, and Swahili, the rest of the Niger-Congo family are almost all tonal. So are the Nilo-Saharan and Khoisan families. Afroasiatic languages, other than Semitic (e.g. Arabic and Amharic) and Berber, are also tonal. With a wealth of complicated tonal data, African languages have contributed significantly to the advance of autosegmental phonology.

While only a few Asian languages use tone to distinguish word categories, African languages frequently exploit tone to signify grammatical categories such as tense, gender, number, and so forth. Shona (spoken in Zimbabwe) contrasts lexical and/or grammatical meanings of the following set of words by tone alone: (1a) /vá-nò-pàr-à/ 'they offend', (1b) /và-nó-pàr-à/ 'the ones who offend', (2a) /vá-nò-pár-á/ 'they scrape', (2b) /và-nó-pár-á/ 'the ones who scrape'. Similarly, tone in Hausa (spoken in Niger and northern Nigeria) differentiates lexical meanings, e.g. /góoràa/ 'bamboo' vs. /gòoráa/ 'large gourd', and grammatical meanings, e.g. /màatáa/ 'wife' vs. /máatáa/ 'wives', and /kíráa/ 'call' vs. /kíràa/ 'calling'. Indeed, the grammatical tone predominates in most African languages. Another remarkable phenomenon observed in many African tone languages is 'downstep', in which a floating low tone (resulted from deletion of segments) causes the pitch of a high tone to be lowered.

With typological similarity to Southeast Asian languages, some West African languages have developed syllable-tone systems, e.g. Grebo (spoken in Liberia). The language has four primary tonal categories (the lowered numeral₁ indicates the highest pitch): (A)

TABLE 2 Varying Pitch Values of the Tonal Categories in Major Dialects of Mandarin

Tonal Categories	Beijing	Wuhan	Guiyang	Chongqing	Kunming
(1) *Yin-ping* (High-level)	55	55	55	55	44
(2) *Yang-ping* (Low-level)	35	213	21	11	31
(3) *Shangsheng* (Rising)	214	42	53	42	53
(4) *Qusheng* (Departing)	51	35	24	24	212

TABLE 3 Tonal Categories in the Pitch-Accent System of Prinmi (R=Rising)

Category	Parameters	Monosyllable	Disyllable	Trisyllable	Quadrisyllable
A	[1] [−spread]	Falling	**H**-L	**H**-L-L	**H**-L-L-L
B	[1] [+spread]	High	**H**-H	**H**-H-L	**H**-H-L-L
C	[2] [−spread]	Rising	L-**H**	L-**H**-L	L-**H**-L-L
D	[2] [+spread]		L-**H**	L-**H**-H	L-**H**-H-L
E	[3] [−spread]		L-R	L-L-**H**	L-L-**H**-L
F	[3] [+spread]			L-L-**H**	L-L-**H**-H
G	[4] [spread]			L-L-R	L-L-L-**H**

High Rising: /na$_{21}$/ 'drink', (B) High: /na$_{22}$/ 'fire', (C) Mid Rising: /hɛ̃$_{32}$/ 'four', and (D) Low: /tɛ̃$_{44}$/ 'test'.

Mende (spoken in Sierra Leone) has a word-tone system with five major tonal categories: H, L, HL, LH, and LHL. The following words belong to the final category: /mbã/ 'companion', /nyàhâ/ 'woman', and /nìḱinì/ 'groundnut'. Etung (spoken in Nigeria) also appears to be a word-tone language. It has eight tonal categories: H, L, HL, LH, HLH, HHL, LHL, and LLH.

Somali (spoken mainly in Somalia, Ethiopia, and Kenya) is a pitch-accent language. According to Saeed's *Somali* (1999), its tonal system is organized into three major categories: (A) H on the last mora (a short vowel has a mora, whereas a long vowel or diphthong has two), (B) H on the penultimate mora, and (C) H on nowhere, i.e. L throughout. Note that the locus of H in the templet is counted from the right. Furthermore, Somali does not multiple its tonal categories with H-spreading; H always spreads leftward to the vowel in the same syllable. The three tonal categories are exemplified as follows: (A) /wéyl/ 'female calf', /ˋinán/ 'girl'; (B) /wéỳl/ 'male calf', /ˋinàn/ 'boy'; and (C) /àmà/ 'or', /kùlùlàà/ 'hot (past tense)'.

Although representative syllable-tone, word-tone, and pitch-accent languages can be identified, a large number of African languages do not fit into these prototypes; many have been regarded instead as 'restricted tone languages'. Some of them might have pitch-accent systems. Many Bantu languages appear to be intermediate between word-tone and pitch-accent systems. We shall arrive at a better classification of these languages when a thorough understanding of the fundamental organization of their tonal systems is available.

Other Tone Languages

Many indigenous languages of the Americas, especially those of Central America, are tonal. Tone development differs considerably among North American languages, e.g. while Achumawi (of northern California) has tone, no tone contrast exists in Atsugewi, a closely related language. Like African languages, American languages employ tone for both lexical and grammatical contrasts.

The San Juan Copala dialect of Trique (spoken in Mexico) shows a syllable-tone system with eight tonal categories, e.g. /yã$_{21}$/ 'be sitting', /yã$_{32}$/ 'corncob', /yã$_{33}$/ 'he is sitting', /yã$_{34}$/ 'salt', /yã$_{35}$/ 'scar', /yã$_{44}$/ 'unmarried', /yã$_{55}$/ 'one (in certain number phrases)', and /yã$_{53}$/ 'Spanish moss'. Similarly, minimal sets of tonal contrast in Sarcee (spoken in Canada) may distinguish word meanings or grammatical meanings: /míɬ/ 'moth' vs. /mīɬ/ 'snare' vs. /mìɬ/ 'sleep' and /-ló/ 'hurt (continuative)' vs. /-lō/ 'hurt (imperfective)' vs. /-lò/ 'hurt (perfective)'. Stress plays an important role in the tonal system of some Central American languages such as the Chinantec languages.

Donohue reports in *Linguistic typology* (1997) that many languages in New Guinea have tone. Iau (spoken in Irian Jaya, Indonesia) has a syllable-tone system with eight tonal categories: /du^{44}/ 'wild pig', /u^{33}/ 'tree', /da^{45}/ 'dog', /tai^{23}/ 'landed', /tai^{42}/ 'fallen', /tai^{423}/ 'falling', /of^{43}/ 'arm', and /ko^{32}/ 'breadfruit'. Teléfól (spoken in Papua New Guinea) is a simple syllable-tone language with a rising tone against a falling one: /kŭl/ 'hand' vs. /kûl/ 'frog'. Word-tone languages include Enga, Kairi, and Usarufa, etc. Kairi has four tonal categories: H, LH, HL, and LHL. The final one always contains a falling tone, e.g. /pâ/ 'tree species', /tùĥi/ 'lid', and /àràmèrâu/ 'basket type'. Pitch-accent languages are numerous in New Guinea, but tone-spreading is generally absent. Una (spoken in Irian Jaya) generates four tonal categories by placing the marked H on the *n*th syllable, where *n* varies from 0 to 3: (A) /kàl/ 'marsupial species'; (B) /kál/ 'tree species', /kúnkàlyà/ 'joint'; (C) /kùnkályà/ 'pimple'; and (D) /bìlbìlýi/ 'building materials'.

References

Bogers, Koen, Harry van der Hulst, and Maarten Mous (eds.) 1986. The phonological representation of suprasegmentals: studies on African languages offered to John M. Stewart on his 60th birthday. Dordrecht and Cinnaminson, NJ: Foris Publications.

Comrie, Bernard (ed.) 1987. The World's major languages. Oxford and New York: Oxford University Press.

Comrie, Bernard, Stephen Matthews, and Maria Polinsky (consultant editors) 1996. The Atlas of languages: the origin and

development of languages throughout the World. New York: Facts on File.

Donohue, Mark. 1997. Tone systems in New Guinea. Linguistic Typology 1.

Fox, Anthony. 2000. Prosodic features and prosodic structure: the phonology of suprasegmentals. Oxford and New York: Oxford University Press.

Heine, Bernd, and Derek Nurse (eds.) 2000. African languages: an introduction. Cambridge and New York: Cambridge University Press.

Hulst, Harry van der (ed.) 1999. Word prosodic systems in the languages of Europe. Berlin and New York: Mouton de Gruyter.

Kaji, Shigeki (ed.) 1999. Proceedings of the Symposium—Cross-Linguistic Studies of Tonal Phenomena: Tonogenesis, Typology, and Related Topics. Tokyo: ILCAA, Tokyo University of Foreign Studies.

Mazaudon, Martin. 1977. Tibeto-Burman tonogenetics. Linguistics of the Tibeto-Burman Area 3.

Merrifield, William (ed.) 1977. Studies in Otomanguean phonology. Dallas, TX: Summer Institute of Linguistics.

Odden, David. 1995. Tone: African languages. The handbook of phonological theory, ed. by John Goldsmith. Oxford and Cambridge, MA: Blackwell.

Pike, Eunice. 1992. Analysis of tone systems. Notes on Linguistics 56.

Saeed, John. 1999. Somali. Amsterdam and Philadelphia, PA: Benjamins.

PICUS SIZHI DING

Translation

Translation occurs when a message or text produced in one language (the 'source' language) is converted into a message or text in another language (the 'receptor' or 'target' language). When used in contrast to interpreting (the conversion of spoken text), translation denotes the conversion of written text. Conversion, however, may often be across different modes, as in 'sight translation', in which a written source text is converted into a spoken target text. Here, therefore, we take the wider definition. Messages can also be translated between varieties of one language (e.g. supplying standard-language subtitles for dialect film dialogue).

Translation studies is a long-established discipline that examines both theory and practice. It has three main concerns:

- Prescriptive: To give practical advice to translators. This tradition has existed for millennia: written advice dates back, e.g. to Cicero in ancient Rome, or to Zhi Qian in third century CE China. More recently, it expresses itself in the search for reliable means of assessing translation quality and in an academic interest in translator training issues.
- Philosophical: Scholars and philosophers, from the medieval Jewish cabalists to Jacques Derrida in modern France, have explored how translation, as communication that transcends a single language, might shed light on key issues of human communication, for example, whether there might have been a single ancestor of all the world's languages, whether meaning is tied to language or independent of it, or whether one person's words mean the same when spoken by another.
- (Applied) linguistics: Nowadays, however, translation studies is usually seen as a discipline that explores the nature of cross-language communication. It draws on (and contributes to) a wide range of fields, such as semantics, pragmatics and discourse analysis, psycholinguistics, communication science, and culture and gender studies.

A key concept in translation studies is that of 'equivalence': the degree of similarity between source and target text or text element (section, paragraph, phrase, or word). In size terms, the target may be roughly as long as the source; it may summarize the source (e.g. an English précis of Japanese company documentation); or it may be an exegetic (explanatory) expansion of the source (e.g. when translating holy scriptures). In qualitative terms, the target may

- Parallel the source's exact words and syntax (an 'interlinear' translation), often in order to explain or gloss the source.
- Reproduce the source's semantic content (a literal translation).
- Convey the source's communicative effect (a 'communicative' or 'dynamic-equivalent' translation), e.g. when translating the Chinese *Chi fan le ma?* (literally, 'Have you eaten yet?') as 'How are you?'

Establishing 'correct' equivalence becomes especially problematic when source and target languages are very different in terms of

- Culture (e.g. how can one convey the sense of the Arabic word *jihad* to a reader with no understanding of Islam?).
- Time (e.g. should one translate Shakespeare into modern or Renaissance French?).
- Linguistic explicitness: For instance, classical Chinese poetry's lack of pronouns causes problems for the translator into modern English (should the subject be 'I' or 'she'?).

Recent models of translation see texts within the framework of a wider communicative act, in which the translator—as source-text receiver and target-text producer—is a key actor:

Original writer/speaker → Writes/speaks → Source text → Read/heard by → Translator → Writes/speaks → Target text →Read/heard by → Target reader/listener

An expert translator has reading and writing (or listening and speaking) expertise in one or more specialist genres; moreover, one source text may have several different equivalents, none of which is necessarily the best. Hence, the translator is now seen as less of a reproducer and more of a creator or re-creator, making active choices on the basis of genre knowledge, intuitive feel, and personal preference. At the whole-text level, translators may have to choose between

- Source-oriented translation: Staying faithful to the original writer's or speaker's wording and intentions. This may involve 'foreignizing' (deliberately retaining source-language or source-culture features), 'idiosyncratic' translation (reproducing the original's stylistic individualities), or both.

- Target-oriented translation: Keeping the target text's new purpose and readership uppermost in mind. For example, the great Arab translator Hunayn Ibn Ishaq (809–873 CE) chose to convert the Greek specialist medical manuscripts of Galen into popular-medicine handbooks. This may also involve 'domesticating' (adapting the text to target-culture norms), 'smoothing' (removing the source's stylistic individualities), or both.

In recent years, translation process studies have examined how translators translate. Among the typical features of translating written prose revealed by these studies are the following:

- During translation, the translator reads ahead a sentence or so at a time, analyzing sentences into clauses and clauses into phrases.
- By default, the actual translating tends to occur phrase by phrase.

- Translators regularly backtrack to modify the emerging target text in terms of word choice, discourse, style, and so on.
- First drafts are usually revised, preferably at a later date.

In much technical or commercial translating (as opposed to literature or advertising, for example), there may be no conflict between source- and target-oriented tactics, because many source and target genres (e.g. Dutch and American air-conditioning documentation) follow the same textual rules. At word, phrase, or clause level, however, problems inevitably occur; much of the translator's creativity consists of selecting strategies for solving such problems. Among the translation strategies identified by translation process research are the following:

- Consulting a dictionary or other reference materials.
- Paraphrasing or adding an explanation.
- Compensation (e.g. replacing a pun by an idiom, or by a pun in a different place).

The translator may also have to make ideological decisions. Choosing whom to translate and not to translate in a region of conflicting identities (e.g. 1990s Bosnia and Serbia) inevitably means taking sides. Translating into a world language such as English, especially if domesticating tactics (see above) are used, may risk contributing to a neoimperialist McDonaldization of world culture, although it may equally be a way for 'minor' cultures to make themselves heard on their own terms. Feminist translation theory raises ideological issues that extend beyond gender, such as whether translators should 'improve' offensive (e.g. sexist) texts or the role of the translator in rediscovering forgotten (female) writers.

References

Baker, Mona. 1992. In other words: a coursebook on translation. London: Routledge.

Baker, Mona. (ed.) 1998. The Routledge encyclopedia of translation studies. London: Routledge

Chan, Sin-Wai, and David Pollard (eds.) 1995. An encyclopaedia of translation: Chinese-English, English-Chinese. Hong Kong: Chinese University Press.

Fawcett, Peter. 1997. Translation and language. Manchester: St. Jerome.

Steiner, George. 1992. After Babel: aspects of language and translation, 2nd edition. Oxford: Oxford University Press.

Venuti, Lawrence. 2000. The translation studies reader. New York: Routledge.

Von Flotow, Luise. 1997. Translation and gender. Manchester: St. Jerome.

FRANCIS R. JONES

Trubetzkoy, Nikolai Sergeyevich

Nikolai Sergeyevich Trubetzkoy was an important member of the Linguistic Circle of Prague, a group of linguists interested mainly in the analysis of the sound systems of languages. Following the French linguist Ferdinand de Saussure, these linguists based their work on a distinction between parole, language being used, and langue, language as knowledge. In terms of sound systems, this distinction appears as a clear division between phonetics, the science of speech sounds as actually produced by speakers, and phonology, the science of language sounds as mental representations.

Trubetzkoy published extensively during the 1920s and his writing defines the Prague School views on phonology. The most comprehensive and explicit formulation of his theory can be found in his *Grundzüge der Phonologie*, published posthumously in 1939 (English translation 1939, *Principles of phonology*).

Phonetics, as Trubetzkoy sees it, is the study of speech sounds as they are physically instantiated, i.e. it deals with the production or articulation of speech sounds and their acoustic features that are relevant for perception. Phonology is a level of analysis that is one step removed from the physical dimension: it abstracts away from the variability of actual pronunciations and attempts to filter out the constant characteristics or features of speech sounds that, for example, allow a hearer to perceive equivalent utterances of different speakers as essentially the same. Thus, phonology describes the abstract mental representations of speech sounds that must be more or less the same for all speakers belonging to a given speech community. While the one is more concrete and the other more abstract, phonetics and phonology nevertheless inform on and borrow concepts from each other.

A phonological description starts with the discovery of those phonetic features that differentiate meanings in a given language. Pit and bit mean different things in English and differ only insofar as the first sound bit is produced with vibrating vocal cords ([b] is 'voiced'), while the first sound in pit is not ([p] is 'voiceless'). Thus, the voiced/voiceless distinction is crucial in English (it is not necessarily so in other languages), and voicing is thus referred to as a 'distinctive feature' in the phonology of English. *Principles of phonology* concentrates mainly on such distinctive features.

Building on a concept elaborated by the Russian linguist Jan Baudouin de Courtenay, Trubetzkoy argued that words contain 'phonemes,' basic sound units that distinguish one word from another. Trubetzkoy was a structuralist thinker, i.e. he did not think that phonemes could be defined by themselves, but only in relation to each other. The basic measure of difference is opposition; a sound stands in opposition to another insofar as the presence of one means the absence of the other. If the presence or absence of a given sound unit creates differences in meaning in a given language, it is a phoneme of that language.

This concept of phoneme developed by Trubetzkoy is typical of the so-called functionalist approach of the Linguistic Circle of Prague. For Trubetzkoy, the phoneme can be defined neither by its psychological character nor by its relationship with its phonetic variants, but by its function. This is where Trubetzkoy's thinking differs from that of his contemporaries such as Leonard Bloomfield or Daniel Jones.

Once the phoneme inventory of a language is established, the phonologist must determine the distinctive features that differentiate the phonemes from one another. /p/ and /b/ differ, for example, in voicing, while /p/ and /t/ differ insofar as they are articulated in specific different locations in the mouth. In his work on distinctive features, Trubetzkoy developed two important concepts: 'neutralization' and 'archiphoneme'. Sometimes, different phonemes of a given language are pronounced exactly alike, i.e. their distinctive feature has been suppressed. If this occurs in a systematic fashion, e.g. in a particular position within a word, Trubetzkoy says that the distinctive feature has been neutralized. German, for example, distinguishes the phonemes /t/ and /d/, i.e. voicing is a distinctive feature just as in English. German, however, devoices sounds at the end of words: *Rad* 'wheel' is pronounced exactly like *Rat* 'counsel'. Thus, German neutralizes the voicing distinction in word-final position. Trubetzkoy calls the representation of such neutralized elements 'archiphonemes'. Thus, an archiphoneme is a phonological unit characterized by only those distinctive features that have not been neutralized. Thus, Trubetzkoy's idea of the archiphoneme is related to the concept of phonemic 'underspecification' in modern distinctive feature theories.

Phonemes and distinctive features are the main phonological elements that Trubetzkoy used to describe the phonological system of a given language. These elements are defined through their relationships with the other elements of the phonological system of that language. These ideas are the basis of the structuralist (and

functionalist) phonology for which Trubetzkoy is mostly known.

However, Trubetzkoy is also well known as the first linguist who called for a new discipline, namely 'morphonology' (or 'morphophonology'), to describe the interaction of phonological and word structures (morphemes). In his principal work on morphonology ('Gedanken über Morphonologie'), Trubetzkoy discusses the phonological structure of morphemes and develops a theory of how the phonological shape of morphemes changes under particular conditions.

Biography

Nikalai Sergeyerich trubetzkoy was born in Moscow, Russia on April 4, 1890. He began studying Russian ethnography and folklore at the age of 13 and published on Slavic languages and Finnish folklore before entering university. He studied at the University of Moscow in 1908–1912. He received his doctorate from Moscow in 1913 with a dissertation on the future tense in Indo-European. He became Assistant Professor, University of Moscow in 1915; he fled to Rostov after the 1917 Russian Revolution, where he taught Slavic Philology at the University. He was Professor of Comparative Linguistics at the University of Sofia (Bulgaria), 1920–1922, and Professor of Slavic Philology, University of Vienna, 1922–1938.

He was a member of the Prague Linguistic Circle since 1926 and a member of the Academy of Sciences at Vienna, in 1930. Trubetzkey died on June 25, 1938.

References

Anderson, Stephen R. 1985. Phonology in the twentieth century: theories of rules and theories of representation. Chicago: University of Chicago Press.

Fischer-Jorgensen, Elli. 1975. Trends in phonological theory. Copenhagen: Akademisk Forlag.

Jakobson, Roman. 1939. Nécrologie Nikolaj Sergejevic Trubetzkoy.' Acta Linguistica 1; Reprinted in Sebeok, Thomas A. (ed.) 1966. Portraits of linguists. A biographical source book for the history of western linguistics, 1764–1963. Bloomington: Indiana University Press.

Jakobson, Roman (ed.) 1989. N. S. Trubetzkoy´s letters and notes. Berlin: Mouton [Mostly in Russian].

Trubetzkoy, Nikolai Sergeyevich. 1929. Zur allgemeinen Theorie des phonologischen Vokalsystems. Travaux du Cercle Linguistique de Prague [TCLP], 1.

———. 1931. Gedanken über Morphologie. TCLP, 4.

———. 1933. La phonologie actuelle, Journal de psychologie. 30.

———. 1934. Das morphonologische System der russischen Sprahe. TCLP, 5.

———. 1939. Grundzüge der Phonologie, TCLP, 7 (1939, posthumous). English translation, Principles of Phonology, Los Angeles, CA.

Trnka, Bohumil 1958. 'Prague structural linguistics'. Philologica Praguensia 1.

ELENA BATTANER-MORO

See also **Phonetics; Saussure, Ferdinand de**

Tuareg and Berber Languages

Berber, which is the mother tongue of the first inhabitants of North Africa, is a Hamito-semitic language. It is spoken chiefly in North Africa, namely: Morocco, Algeria, Tunisia, Libya, and Egypt. Berber is also spoken by the Tuareg populations in Mali, Niger, and Burkina-Faso. The regions where Berber is used are discontinuous in nature; that is, they are surrounded by Arabophone or other populations. These discontinuous groups may range in size from a tribe, a village, or even a greatly populated center.

The Berber population can be divided into four major groups. First, we have Morocco, where the Berberophones are estimated to comprise about 40% of the overall population. Second, there is Algeria, in which approximately 25% of the population is Berber (cf. Grandguillaume 1983; Chaker 1990 and Ennaji 1991). This means that the Berber-speaking population is around 9.5 million in Morocco and 5.5 million in Algeria. Third, we have the Tuareg populations, which represent around 3 million people scattered in the sub-Saharan countries, namely: Niger, Mali, and Libya. The fourth population group totals about 140,000 people; these are generally in small and isolated areas, particularly in Tunisia (less than 100,000) Mauritania (about 10,000 people), and Siwa in Egypt (30,000 people). cf. Chaker (1990) and Mustapha (1993).

Tuareg

Tuareg is a variety of Berber spoken in Mali and Niger in West Africa; it has an alphabet called Tifinagh. Tuareg people used to be called 'hommes bleus' (blue people) because they lived in an immense desert and were brave and respectable people. The Tuareg were colonized by France until 1960 when Mali and Niger obtained independence.

Today, Tuareg people are scattered over Mali and Niger; after four decades of independence, Tuareg people living in these two countries feel excluded and their language and culture marginalized. Many cultural associations in the region militate for the official recognition and revitalizing of Tuareg language and culture. In a letter addressed to the United Nations on August 1, 1997, Abdoulahi Attayoub, president of the association 'Survie Touarègue' asked the UN to protect the Tuareg minorities in Mali and Niger whose rights were violated and whose lives were in danger because of racial discrimination instituted by the regimes in place. Thus, thousands of Tuareg people were killed by the army in both countries between 1980 and 1995. Others fled their country and settled in Algeria, Mauritania, and Burkina-Faso, where they still live in hard conditions (see the journal *Tifinagh* 1998, issue 3).

Since the periods of independence, the different governments that took power in Niger and Mali marginalized the Tuareg people and often attempted to wipe out their language and culture. The Tuareg regions have been deprived of all hope of socioeconomic development, and their cultural space has been denied to them.

The Tuareg people live in Central Sahara and on the borders with the Sahel. They call themselves Imajighen, Imoushagh, or Imouhagh depending on the regions; their language is called Tamajeq, Tamasheq, or Tamaheq. The Tuareg language is generally considered the most well-preserved and the richest of the Berber languages. Tuareg people have retained an original alphabet called Tifinagh, which is derived from an ancient Berber writing system that was used in all North Africa and in the Sahara.

The oral tradition in Ahaggar shows that Tuareg people are the descendants of Tin-Hinan, queen and mythical ancestor of the north. Their political system was based on a confederation, where power belonged to the direct descendants of this queen mother, who is the founder of the linguistic community of the Tuareg language. All Tuareg people have a common history, the same language, and the same lifestyle, which are the bases of a solid feeling of belonging to their land and of a common cultural identity called Temoust.

Traditionally, there were seven Tuareg confederations, which are today distributed as follows: Ahaggar (in Algeria and Niger), Ajjer (in Libya and Algeria), Aïr (in Niger), Azawad (in Niger and Mali), Adghagh (in Mali), Tadamakkat (in Mali), and Oudlan (in Burkina-Faso). Today's political borders are inherited from the days of colonization; they have cut the Tuareg country into many parts that are integrated into the five countries cited above.

Because of the lack of a census where the linguistic factor is taken seriously into account, the population of Tuareg people is estimated to be around three million: 1.5 million in Niger, 1 million in Mali, and 500, 000 in the other countries (Algeria, Libya, and Burkina-Faso). Today, the Tuareg people, who are generally proud of their langue, struggle to freely use their mother tongue and their culture. But because of the pressure put by the authorities in both Niger and Mali on Tuareg people, some young native speakers feel it is degrading to speak *Tamacheq* in public (see Elékou 1998).

Berber Languages

Of all the North African countries, Morocco has the largest Berber-speaking population, followed by Algeria. The above estimates are only tentative, as exact figures are impossible to provide given that the old censuses in Morocco (1982) and in Algeria (1986) did not take into account the linguistic situation. However, the most recent Moroccan census (organized in 1994), which was also concerned with the linguistic factor, reveals that approximately 11 million Moroccans speak Berber today.

Nearly half of the Moroccan population is Berberophone; i.e. they are monolingual Berbers, and the majority of them are bilingual because they also speak Dialectal Arabic. The monolinguals are basically very small children and old people who do not participate in active life, whereas the bilinguals are usually the young and the educated. Berberophones are considered illiterate as long as they do not master Standard Arabic or French.

Berber is a spoken language with virtually no real written history. Unlike the North African populations, the Tuareg populations in the south of Algeria and Mauritania — as mentioned above — have preserved an old writing system called *Tifinagh*, which has been handed down from generation to generation through centuries. However, there is no evidence that *Tifinagh* has ever been utilized to write down history or literature.

Today, the Classical Arabic script is at times used by creative writers, close friends, or members of a family to write personal letters. Resorting to writing Berber in Arabic script is often due to the lack of mastery of the Classical Arabic language. The Moroccan Association of Cultural Exchange publishes a periodical in Berber entitled *Amud*, where Berber is written by means of the Arabic writing system. In Algeria, Berber textbooks and novels are written in Latin script (cf. for example, the works of Mouloud Mammeri, Rachid Aliche, and Ramdane Achab). The Tifinagh script, which has been recently revived and modernized, is widespread in the Berber populated area of Kabylie (see Chaker 1984:37).

In Morocco, there are three main Berber dialects. First, there is Tarifit used in the Rif, north of the country. Second, there is Tamazight, spoken in the Middle Atlas, and east of the High Atlas. Third, there is Tashlhit, spoken in the High Atlas and the Anti-Atlas Mountains, south of Morocco.

The Berber dialects spoken in Morocco are generally mutually intelligible unless the two varieties in question are situated at extreme geographical points. For example, a native speaker of Tarifit in the north can hardly understand or communicate with a speaker of Tashlhit in the south. Additionally, the Berber dialects used in Morocco and Algeria are usually mutually incomprehensible.

In Algeria, there exist four major Berber dialects. First, there is Kabyle, which is spoken in the area of Kabylie, near Algiers and Bougie. Second, there is the Shawia dialect spoken by the populations of Aures, north of Constantine. Third, there is the Tuareg dialect, which is a variety that is largely used in the extreme south of Algeria. Fourth, there is the Mzab dialect in the Ibadhits, notably in Ghardaia.

In Tunisia, the Berber population is less important. Berber is spoken only in the extreme south of the country. There exist a handful Berberophone villages, chiefly situated in Djerba, in the west of Matmata, and in the east of Gafsa (cf. Payne 1983; Chaker 1984).

The traditional segmentation of Berberophone regions into zones does not genuinely reflect the considerable effect of rural exodus that the Maghreb (Morocco, Algeria, Tunisia, and Libya) has witnessed since Independence. As a result, there are nowadays vast Berberophone populations in the urban centers of the Maghreb. Thus, Berber is also spoken in the cities of Algiers, Oran, Tizi-Ouzou, Agadir, Casablanca, Rabat, Fès, Nador, Tangier, and, in fact, in most large cities (see Chaker 1990; Ennaji 1991).

Although there are many Berber dialects, Berber is fundamentally one language (cf. Boukous 1975; Chaker 1984). Although most Berber dialects are generally mutually unintelligible, they share similarities at the phonological, morphological, syntactic, and lexical levels, as will be seen below (cf. Basset 1959; Applegate 1970 for details).

The Linguistic Features of Berber

Historically, Berber has borrowed mainly from Latin, Arabic, and French. The remnants of Latin loanwords are limited to a dozen or so. Nonetheless, the most well-known loans come from Dialectal Arabic and French. Most of these borrowings are nowadays completely adapted to the patterns of Berber. For instance, the Latin words *asinus* (little donkey), *burrhus* (coat), *tussis* (cough) are pronounced *asinus, abernus,* and *tusut* in Tamazight, the Berber dialect of the Moroccan Middle Atlas; The Dialectal Arabic words *sa:'a* (watch) and *Tbib* (doctor) become *tassa't* and *adbib*, respectively, in Berber. The French loans *veste* (vest) and *automobile* (car) become *lfista* and *Tomobil*, respectively, in Berber (see Ennaji 1991).

On the other hand, Berber itself has affected Dialectal Arabic in many ways. A multitude of Berber terms are incorporated in the latter language. If we take the Moroccan context for illustration, we notice that Berber intonation has infiltrated the Moroccan Arabic sound pattern, especially in the cities neighboring Berber villages. Moroccan Arabic is also full of Berber loans; for instance, the following Moroccan Arabic words originate from Berber: *brra* (outside), *berrm* (turn), *atay* (tea), *taRunja* (prayer for rain), and *tiwizi* (collective land ploughing).

Most of these lexical items are fully integrated in the Dialectal Arabic inflectional system and so widespread that few Moroccan Arabic speakers today will recognize them as Berber. However, the more a Berber dialect is geographically distant or isolated from the urban centers, the less affected it is by the Arabic language. This is evidenced by the fact that in towns and cities, Berberophones tend to use more Arabic loans than in the rural Berber areas. Nonetheless, the young generation of Berberophones is much more influenced by Arabic than the old generation. This is mainly due to the importance of urbanization and the expansion of schooling and Arabization in the educational system (see Ennaji 1988; Elbiad 1991).

Linguistically, Berber has the following intrinsic features. First, phonologically, it resembles Dialectal Arabic in that most of the consonants and vowels existing in Dialectal Arabic do exist in Berber. Nevertheless, there is a difference as far as the verb system is concerned. The latter is not much affected by Dialectal Arabic. For example, most of the loan verbs are integrated in Berber verb morphology with no consequence on the verb group pattern. By way of illustration: *xdm* (work) --> *i-xdm* (he worked); *jawb* (answer) ---> *t-jawb* (she answered). This shows that the Berber verb inflection is essential for sentence formation and meaning, which accounts for the verb system's resistance to Arabic influence.

Evolution of Berber

Berber has experienced a gradual regression in Tunisia, Morocco, and Algeria. As we have already mentioned, the number of Berber speakers in Tunisia has dramatically dwindled to about 100,000, while in Morocco and Algeria the number of competent native speakers of Berber is also regressing, especially in urban areas where Dialectal Arabic is more powerful and commonly used.

Four major reasons lie behind the regression of Berber. First, there is the phenomenon of rapid urbanization, which is itself the result of migration from rural areas to urban centers. Thousands of Moroccan Berberophones have quit their homes and villages in the last four decades to settle down in cities

like Casablanca, Agadir, Rabat, Nador, El-Hoceima, Fès, and Marrakesh to better their lives.

Second, there is the expansion of free education since Independence. Thus, the schooling of Berber-speaking children meant their Arabization and their gradual loss of competence in Berber for the latter is excluded from education. This is a factor that can be remedied if Berber is seriously taught as a language in public schools.

Third, there is the fact that most Moroccans are aware that Berber is pragmatically a local language whose use is not widespread, unlike Arabic or French, and hence it is not useful in business, science, technology, and international affairs. Its role is merely confined to the cultural, social, and historical domains (cf. Ennaji 1991).

Finally, the fact that Berber is officially treated as 'a dialect' discourages people from using it in all walks of life. Many Berberophone parents no longer interact with their own children in Berber, simply because they feel that Berber lacks prestige and does not enjoy the status of a language (see Boukous 1997). One way of verifying the regression of Berber is to examine the domains in which it is normally used. Unlike in the 1940s and 1950s when Berber was spoken not only at home but also in the public administration and in political circles, nowadays Berber is virtually limited to home and to conversation among close friends. Dialectal Arabic is rapidly gaining ground to the detriment of Berber even in homes of Berberophones and even in rural areas where Berber was hitherto the predominant language of communication. For instance, Berber is implicitly rejected in administrative offices, in the private sector, and in political spheres (cf. Sadiqi 1997; Boukous 1997). Because of this rapid decline, intellectuals and progressive forces in the region have recently begun a campaign seeking the revitalization of the Berber language and culture. This campaign has resulted in positive changes in officials' and people's attitudes toward Berber, as we will see in the section below.

Tuareg and the other Berber languages are spoken in the Maghreb and the Sahel. They enjoy both historicity and vitality. They are historically among the oldest languages in the world. Tuareg Berber is marginalized although it is spoken by approximately 3 million people scattered around five different countries: Mali, Niger, Burkina-Faso, Algeria, and Libya. It is the Berber cultural influence that gives Morocco and Algeria their history, and that makes them so different from the rest of the Arab world. Berber remains vital because it is still spoken by a great part of the population in North Africa. The recent revival of interest in Berber language and culture is a good indication that its vitality will be strengthened to the extent that it may be raised to the level of a standardized and codified language in the future.

References

Applegate, J.R. 1970. The Berber languages. Current Trends in Linguistics 6. 586–61.

Basset, André. 1959. Articles de dialectologie berbère. Paris: Klinksiek.

Boukous, Ahmed. 1995. La langue berbère: maintien et changement. International Journal of the Sociology of Language 112. 9–28.

Boukous, Ahmed. 1997. Situation Sociolinguistique de l'Amazighe. International Journal of the Sociology of Language 123. 41–60.

Chaker, Salem. 1984. Textes en linguistique berbère. Paris: CNRS.

Chaker, Salem 1990. Langue berbère: une planification linguistique extra-institutionnelle. Maghreb linguistics, ed. by J. Pleines. 237–63.

Elbiad, Mohamed. 1991. The role of some populations sectors in the progress of Arabization in Morocco. International Journal of the Sociology of Language 87. 27–44.

Elékou, Moussa, Ag. 1998. Les Touareg au Niger: Repères Historiques. Tifinagh 13. 95–7.

Ennaji, Moha. 1988. Language planning in Morocco and changes in Arabic. International Journal of the Sociology of Language 74. 9–39.

Ennaji, Moha. 1991. Multilingualism in the Maghreb. International Journal of the Sociology of Language 87. 7–25.

Grandguillaume, Gilbert. 1983. Arabisation et politique linguistique au Maghreb. Paris: Maisonneuve.

Mustapha, A.A. 1993. Siwa: a fight between death and life (in Arabic). Al-Arabi 411. 134–51.

Payne, R.M. (ed.) 1983. Language in Tunisia. Tunis: Bourguiba Institute of Modern Languages.

Sadiqi, Fatima. 1997. Grammaire du Berbère. Paris: L'Harmattan.

MOHA ENNAJI

Tungusic

The Tungusic languages are spoken by small, mostly rural populations scattered across East Asia. The northernmost language, Even, is found in the northeastern region of Siberia between the Lena and Anadyr' rivers, although one Even speech community is located farther east on the Kamchatka peninsula. The southern

geographic boundary of the Tungusic languages is defined by Solon and Orochen, which are situated in and around the Hinggan mountains of China. To the west, one finds Evenki speakers along the Yenisey River in Russia and Sibe communities in Xinjiang Province in China. To the east, speakers of Orok and Evenki inhabit Sakhalin Island. Despite the large area over which the Tungusic languages are spread, the highest density of Tungusic languages is within the Amur River basin in southeastern Siberia, where there are speakers of Oroch, Negidal, Evenki, Nanai, Udege, Orok, Kili, and Ulch.

Although scholars agree that the Tungusic languages constitute a distinct language family, some uncertainty surrounds exactly how individual languages are related. The family is commonly divided into three main branches: Southern Tungusic (comprising Manchu, Sibe, and the dead language Jurchen), Central Tungusic (comprising Kili, Nanai, Oroch, Orok, Udege, and Ulch), and Northern Tungusic (comprising Even, Evenki, Negidal, Orochen, and Solon). Two points of contention surround this classification. First, disagreements have persisted about how to classify certain languages and dialects, particularly those listed above as Central Tungusic. Second, some linguists argue that Central Tungusic is sufficiently close to Southern Tungusic to warrant combining these two branches into one. Indeed, this notion also characterized many of the early efforts to classify the Tungusic languages, notably the pioneering work of V.I. Cincius and Johannes Benzing. This two-way classification of Tungusic accounts for why the family is commonly referred to as Manchu-Tungusic, with Manchu representing the Southern/Central branch and Tungusic representing the Northern branch.

Regardless of whether a two-way or a three-way classification of Tungusic languages ultimately proves correct, there are certain criteria that have played, and will continue to play, a crucial role in determining the historical relationships among these languages. For example, most Tungusic experts place great significance on the correspondence of word-initial *f-* in Manchu and Sibe with word-initial *p-* in many (but not all) of the Central Tungusic languages and word-initial *h-*, *x-*, or no consonant in Northern Tungusic (e.g. Manchu *faxun* 'liver', Nanai *pa*, Orok *paka*, Negidal *xaxin*, Evenki *hakin*, Orochen *axin*). Another useful characteristic commonly examined for classification purposes is the presence vs. absence of *-g-* between vowels. Where many (but not all) Northern Tungusic languages use the sound intervocalically, it tends not to appear in Southern and Central Tungusic (cf. Orochen *buga* 'land' vs. Manchu *bua* vs. Ulch *bua*).

Traditionally, speakers of most Tungusic languages were organized into small nomadic groups that engaged in hunting or reindeer herding. Their mode of subsistence made it necessary for individual clans to be small and for the clans to be spatially dispersed so as not to overgraze or overhunt particular regions. Although individual clans would remain in contact with each other, the degree of interaction was limited; yet it was not uncommon for individual or families to switch from one clan to another, particularly for the purpose of marriage. The linguistic consequence of these social characteristics was extensive dialect differentiation. These dialects frequently form a dialect chain: speech varieties that are close to each other along the chain are mutually intelligible and are clearly dialect-like in their differences. Varieties farther apart on the chain, however, cease to be mutually intelligible and are thus different enough to be called separate languages. As a result, it is a matter of some difficulty today to make a clear demarcation between languages and dialects; thus, lists of Tungusic languages can vary in content from as few as six members to as many as 17.

An added complication is that Tungusic groups are commonly referred to by names that have been given to them by political authorities and other outsiders. These names do not always accurately reflect linguistic realities. For example, Ewenki (Chinese *Ewenke*) and Orochen (Chinese *Elunchun*) are the names given to two of China's official minorities. The former suggests a connection to the Evenki, one of the official minorities of the Russian Federation. However, some of the Orochen groups in China have a closer linguistic connection to the Russian Evenki than do the Ewenki. Even more confusing is that what the Chinese have designated as a single unit, the Ewenki, turns out to be three linguistically distinct groups, the Solon, the Khamnigan Evenki, and the Yakut Evenki. The last of these groups is intimately connected to the Evenki of Russia, the first two far less so.

Although some points of uncertainty remain regarding the internal classification of Tungusic, much greater debate has emerged over the relationship of the Tungusic family to other language families. In particular, many scholars argue that Tungusic is a member of a macrofamily called Altaic, along with at least Turkic and Mongolic, and perhaps Japanese, Korean, and Ainu as well. Despite a superficial similarity in the grammars of all these languages and suggestive evidence based on shared vocabulary and sound correspondences, proponents of Altaic have not yet been able to demonstrate their hypothesis to the satisfaction of many linguists. Therefore, the question of the connection between Tungusic and other language families remains an exciting area of current research.

The most familiar Tungusic language is Manchu, which served as the language of the Qing dynasty of China (1644–1911), one of the largest empires in his-

tory. Manchu and Jurchen hold a special place within Tungusic as the only two languages for which we have written records that predate the twentieth century. The system used to write Manchu can be described as either an alphabet or a syllabary, i.e. it contained symbols both for individual sounds and for syllables. It was derived from Mongolic script at the very end of the sixteenth century, and the language was used for all manner of official documents and for the translation of many Chinese classics; consequently, a large corpus of written Manchu exists, which has proved invaluable for the reconstruction of proto-Tungusic, the ancestor language, and for attempts to link Tungusic to other language families.

During the reign of the Qing emperors, the use of Manchu steadily declined as more and more ethnic Manchus employed Chinese for their everyday activities. By the time the Qing empire was overthrown in the early twentieth century, the Manchu people had been so thoroughly assimilated into the Chinese language and culture that a linguistically distinct Manchu people had ceased to exist. Despite nearly 10 million people claiming Manchu ethnicity in China today, the best estimates of fluent speakers of the language place the number between five and 30.

The oldest records of a Tungusic language come from Jurchen, a language spoken by the founders of the Jin dynasty of China (1115–1234). These records take the form of inscriptions found in northeastern China and in Korea. The Jurchen writing system, which appears to combine symbols for words as well as for syllables, has not been fully deciphered. However, enough evidence from Chinese lists of Jurchen words exists to establish that the language is closely related to Manchu, either as a historically earlier variety of Manchu or as a dialect parallel to Manchu.

In terms of the number of speakers, the largest Tungusic language is Sibe (Chinese *Xibe*), which, according to the 1982 Chinese census, had 26,760 speakers. Sibe is a dialect of Manchu but has been designated separately within China, primarily because of the presence of a Sibe group in the Jungaria region of Xinjiang Province. Over time this group, which has its origins in Manchu military units sent to protect the frontiers of the Qing dynasty in the mid-eighteenth century, has developed independently from other ethnic Manchu groups. Moreover, the Sibe have been linguistically and culturally influenced by Mongolic and Turkic groups living in Xinjiang.

The only other Tungusic languages that currently have a critical mass of fluent speakers are Solon (estimated 14,000 speakers), Siberian Evenki (estimated 10,000), Even (estimated 7,400), and Nanai (estimated 6,000). In very few of the communities where these language are spoken, however, do children learn them.

Therefore, it is likely that all the Tungusic languages, with the exception of Sibe, will become moribund within the next generation.

In terms of their structural features, Tungusic languages tend to organize their basic clause structure around the pattern Subject–Object–Verb, although individual languages vary in how rigid this word order is. They have large numbers of suffixes, which are used not only on verbs, nouns, and adjectives, but on closed class items such as numerals. For example, Oroqen builds a number of words from numeral roots: *ilan* 'three', *ila-Nna* 'three animals', *ila-la* 'three days', *ila-na* 'a group consisting of three animals', *ila-tal* 'three each', *ila-ra* 'three times', *ila-kan* 'only three', and *ila-ki* 'third'.

Another common characteristic of Tungusic is vowel harmony. Most significantly, the vowels in certain suffixes must match the properties of vowels that are in the word root. For example, in Sibe many derivational and inflectional suffixes have two forms, one with the vowel [u] and one with the vowel [ɨ]. The choice of suffix form is determined by the last vowel in the root. If this root vowel is pronounced with rounded lips, then the suffix containing [u], also a rounded vowel, is employed. If the last root vowel is not rounded, the suffix containing [I], a non-rounded vowel, is used (e.g. *batur* 'hero' yields *batur-lu* 'to be heroic', while *ɔldʒi* 'a capturing' becomes *ɔldʒi-lɨ* 'to capture someone'). Although all Tungusic languages exhibit some harmonizing processes, there is tremendous variation in the form that the harmony takes and in how productive it is. Sibe possesses only rounding harmony, as just illustrated. Other Tungusic languages also have harmony based on the feature of vowel tenseness (e.g. Nanai, where suffixes with a tense vowel, such as the adjective-forming *-su*, as in *densi-* 'to care for'/*d´nsi-su* 'careful', have an alternative form that contains a lax vowel *-sˑ*, as in *sa-* 'to know'/*sa-sˑ* 'well-informed'). Still other Tungusic languages, such as dialects of Oroqen and Solon, have both types of harmony.

Tungusic languages make frequent use of converbs, which are verb forms used for subordinate structures. Consider the following example from Evenki: *si: ↔m↔-tΣ↔l↔ ⫟-s nuŋartin ŋ↔n↔-dʒ↔ŋ↔ ⫟-tin* 'After you come, they will go.' (literally: 'you come they go'). The converb is *↔m↔tʃ↔l↔ ⫟s*, which has been created by attaching *-tʃ↔l↔* to the verb root *↔m↔-* 'come'. This particular suffix indicates that the event expressed by the converb preceded the event expressed by the main verb. Evenki possesses 12 other converb-forming suffixes. The suffixes allow speakers to express many different meanings including simultaneity, posteriority, condition, and purpose.

Another characteristic property of Tungusic languages is that they employ case suffixes on nouns to

indicate grammatical functions. In this regard, the Northern Tungusic languages tend to have far more case inflections than the Southern Tungusic languages. For example, Even differentiates at least 12 cases (*etiken* 'old man', *etike-m* 'old man' (object), *etiken-du* 'old man' (indirect object), *etiken-n'* 'with the old man', *etike-n'un*, *etiken-duk*, *etiken-dula* 'at the old man', *etiken-duli*, *etiken-taki* 'to the old man', *etike-ŋitʃ*, *etike-kla*, *etike-kli*), whereas written Manchu has only five (*ama* 'father', *ama-i* 'of the father', *ama de* 'father' (indirect object), *ama be* 'father' (direct object), and *ama ci*).

Unlike Southern Tungusic, Northern Tungusic languages also contain a set of nominal suffixes that indicate possession (e.g. Oroqen *murin-iw* 'horse-my', *murin-iy* 'horse-your (sing.)', *murin-in* 'horse-his/her', *muri-mun* 'horse-our/excluding the listener', *murin-tir* 'horse-our/including the listener', *murin-ʃun* 'horse-your (pl.)', *murin-tin* 'horse-their').

References

Bulatova, Nadezhda, and Lenore A. Grenoble. 1999. Evenki. München: Lincom Europa.

Doerfer, Gerhard. 1978. Classification problems of Tungus. Tungusica, Vol. 1, ed. by Gerhard Doerfer and Michael Weiers. Wiesbaden: Otto Harrassowitz.

Janhunen, Juha. 1991. Material on Manchurian Khamnigan Evenki. Castrenianumin toimitteita 40. Helsinki: University of Helsinki.

Li, Bing. 1996. Tungusic vowel harmony: description and analysis. Amsterdam: University of Amsterdam.

Li, Gertraude Roth. 2000. Manchu: a textbook for reading documents. Honolulu: University of Hawaii.

Malchukov, Andrei L. 1995. Even. München: Lincom Europa.

Nedjalkov, Igor V. 1997. Evenki. London: Routledge.

Norman, Jerry. 1974. A sketch of Sibe morphology. Central Asiatic Journal 18.

Whaley, Lindsay J, Lenore A. Grenoble, and Fengxiang Li. 1999. Tungusic classification from the bottom up. Language 75.

Zhang, Yan-chang, Bing Li, and Zhang Zi. 1989. The Oroqen language. Changchun: Jilin University Press.

LINDSAY WHALEY

Turing, Alan

Alan Turing made foundational contributions to philosophy, mathematics, biology, artificial intelligence, and computer science. He, as much as anyone, invented and showed how to program the digital electronic computer. From September 1939, his work on computation was war-driven and practical. He developed high-speed computing devices needed to decipher German Enigma Machine messages to and from U-boats, countering the most serious threat by far to Britain's survival during World War II.

Because of official secrecy, his war time exploits were unknown until the 1980s. By then some of his inventions no longer seemed connected to a real human being. For example, Turing's 1936 paper, 'On computable numbers', was soon seen as the most important theoretical paper ever written on computation. Hence, mathematicians, engineers, and computer scientists came to write of 'turing machines' and 'universal turing machines' almost forgetting that there was an Alan Turing.

In his 1936 paper, Turing answered the deepest computational question — whether there is a finite mechanical procedure for deciding whether any given mathematical statement is true or false. Turing realized that the key was to specify what a 'mechanical procedure' (or computation) was. Mathematicians tended to assume that it meant 'explicit, step-by-step, requiring-no-creativity' and left it at that. Turing saw that if a procedure were mechanical, it could be automated.

Turing's answer is to imagine a starkly minimal machine. He called it a 'theoretical computing engine', because he had no intention of building one. This is a machine with which to think. It is fed by an indefinitely long tape divided into frames like those on a roll of film. The machine has a 'read head', which can tell whether the frame under it has a '/', a '\', or is blank. The read head is also a 'write head' that can erase, write in a '/', or '\', or do nothing. The machine can also move one frame forward or backward or stay in place. Then there is another move and so on. At the beginning of each move, the machine is in one of a small number of 'internal states' and this may switch to another after the move. The machine is built to enact instructions in its machine table of the form 'if / is read and the internal state is 1, then erase, move one frame forward, and go into state 2'. A turing machine for adding would get, say, the input sequence '// //' and then automatically change it into '////' through a long series of steps and then stop. For us this is the computation '2+2=4'. Because Turing was thinking about a theoretical device, he did not mind that a million

would be represented by a like number of '/'s. Indeed, he insisted that an actual and literal universal turing machine would be much too slow for practical computation.

Turing goes on to show that anything mathematicians call a mechanical problem or computation can be represented by some turing machine. Further, he shows that there is a universal turing machine. Depending on the input tape sequence, this turing machine can turn itself into any particular turing machine, do a computation, and then turn itself into another turing machine and do a computation, and so on. In 1936, Turing thought up the general-purpose digital computer and gave its definitive abstract description. The tape can also store data and programs ('memory', 'skills', 'plans'), represent incoming data ('sensory input'), and issue output instructions ('motor outputs'). Thus, Turing has also given the framework in which to describe any sort of individual thinker. As many scientists have said since the 1960s, each human is, more or less, a universal turing machine and so are our digital electronic computers; for the same reason, we now think of thinking as computing or data-processing. To return to his original question, Turing also showed that no possible computer (digital electric, human, or alien) can decide the truth or falsity of every mathematical formula. Turing broke the species barrier: given his minimalist physical description of computability, anything could be a mind whatever its physical composition.

His 1950 paper, 'Computing machinery and intelligence', set the agenda for cognitive science. From it, scientists and philosophers extracted the goal of writing 'intelligent' computer programs good enough to pass the 'turing test' for simulating human intelligence. Although some computer scientists boasted that they would program a 'passer' within a decade, they have not come close after five decades. Turing also suggested several other ways, aside from programming, that might be used to simulate human intelligence. When he did so, he anticipated approaches that were explored in the 1980s and 1990s — training connectionist nets, sending a 'child machine' to school, perhaps equipping it with eyes, ears, and hands. The centrality of natural language in Turing's test, and in human intelligence, is reinforced by the continuing failure of machine translation from one human language to another. Computers now easily exceed the most talented humans in arithmetic calculation and chess playing, but none come anywhere close to the performance of ordinary human translators (in merely taking dictation, our most sophisticated voice recognition programs do not come close to matching the performance of ordinary human secretaries).

In his 1950 paper that created the field of artificial intelligence, Turing asks you to imagine the following 'imitation game'. We have some judges who communicate by a terminal to A and B terminals. One of these terminals is operated by a woman, one by a man. Under the judges' questioning, the woman tries to convince them that she is the woman, while the man tries to convince them that he is the woman. For example, the judges might ask, 'How do you do your hair?' Turing comments that the best strategy for the woman is to tell the truth. Interrogators look for inconsistencies. Telling the truth is the simplest way to avoid them. The man is going to have a lot to keep track of. If the man manages to win, you might say that he can think like a woman.

Turing then proposes that we substitute a computer for the man. If the computer 'passes', that proves that the computer can think like a human being and that proves that the computer can think period. Turing remarked that the test draws a clear line between the physical and mental. He likened putting the contestants in separate rooms, so the judges cannot see them, to musical contests where the players perform behind a screen, so the judges would not be biased by their physical appearance. This is what we now call the 'Turing test'. Turing did not think it was going to be easy to create a passer. He provided a vivid and testable goal for artificial intelligence research.

On the other hand, some of Turing's work was so far ahead of his time that it earned credit only in retrospect. His last published paper (1952) anticipated the most important new approach—chaotic, structure-spawning reactions—in the last half century of developmental biology. But it was not until the late 1970s that scientists began to refer to 'turing structures'.

Biography

Alan Turing was born on June 23, 1912, in London. He studied in Sherborne School, 1926–1931 and Wrangler, Mathematics Tripos, King's College, Cambridge, 1931. He received his Ph.D. from Princeton University in 1938; he was Fellow, King's College, 1935–1945; Princeton University, 1936–1938; British Foreign Office, Bletchley Park, 1939–1945; National Physical Laboratory, 1945–1948; and University of Manchester, 1948–1954. He received the Smith's Prize from Cambridge University in 1936 and the Order of the British Empire in 1946. He was Fellow, Royal Society in 1951. In 1952, Alan Turing was convicted of 'homosexual acts' and legally forced to take huge hormone doses that rendered a man who nearly represented Britain in the 1948 Olympic marathon impotent, femininely breasted, and obese. He died of self-administered cyanide on June 7, 1954, in Manchester, England.

References

Hodges, Andrew. 1984. Alan Turing: the enigma. New York: Simon & Schuster.

Leiber, Justin. 1991. An invitation to cognitive science. Oxford: Blackwell.

Turing, Alan. 1936. On computable numbers, with an application to the Entscheidungsproblem. Proceedings of the London Mathematical Society 42.

———. 1947/1970. Intelligent machinery. Machine intelligence, ed. by B. Meltzer and D. Michie. New York: American Elsevier Publishing Co.

———. 1950. Computing machinery and intelligence. Mind. 59.

———. 1952. On the chemical basis of morphogenesis. Philosophical Transactions of the Royal Society of London, Series B. 237. 37–72.

———. 1950. Computing machinery and intelligence. Mind. Also available in many other places including Hofstadter, D., and Dennett, D., The mind's I. New York: Basic Books, 1984.

———. 1952. The chemical basis of morphogenesis. Philosophical Transactions of the Royal Society.

JUSTIN LEIBER

See also **Artificial Intelligence**

Turkey

The name Turkey, as historical Anatolia (and a small part of the European mainland) was officially renamed in 1923, incorrectly implies that the land is home to the Turkish language. In fact, Anatolia was the traditional home of a number of Indo-European languages (notably Ionic Greek in the west, Hittite in the center, Pontic Greek in the north, and Armenian in the east) and several autochthonous languages including Hurrian, Hattic, Urartian, and Laz. Turkish and its closest relatives only began to appear in Anatolia in the eleventh century AD, and during the six centuries of Ottoman rule (thirteenth to twentieth centuries) expanded far beyond the borders of present-day Turkey and largely supplanted its linguistic predecessors. Dialects of Turkish are still spoken in small pockets in the Balkans, especially Greece and Bulgaria, as well as in Syria. More recently several other language families have become established in Turkey, notably Kurdish in the southeast and various northwest Caucasian languages in the north and west.

These minority groups and their languages and cultures were tolerated to some extent during most of the Ottoman period, but the Empire in its last days and the Republic that replaced it (1923–present) has implemented a number of military and legal actions designed to minimize their presence, most importantly the elimination of the sizeable Armenian and Greek populations in 1915–1922, the name substitution program, which replaced more than 30,000 traditional place names (primarily Greek and Armenian) with new Turkish names, and Law 2932, passed in 1983, which prohibits publishing in languages other than Turkish. Children, moreover, are generally forbidden to speak a language other than Turkish at school, and with few exceptions, radio and television broadcasts in minority languages (especially Kurdish) are not allowed. Only the officially recognized religious minorities (Greeks, Armenians, and Jews) are permitted to teach in their respective languages at the primary and secondary school levels. As a result of these constraints and increasing urbanization over the last few decades, the many minority languages of Turkey have been severely eroded, and many have disappeared entirely. There are at least 37 languages currently or formerly spoken natively in Turkey, but most of these are in danger of disappearing in the near future.

The largest language at present is Turkish, spoken by more than 60 million residents of Turkey, about 90% of the population. As of 1995, there were some 13 million additional speakers in other countries, primarily Germany, Bulgaria, Greece, Macedonia and Yugoslavia, the Netherlands, France, Romania, and Cyprus. Other Turkic languages are represented in Turkey as well, including (as of 1982) Azeri (530,000 in the Kars province), Gagauz (327,000), Kazakh (600+), Kirghiz (1,137 in the provinces of Van and Kars), Tatar (several thousand, including the Crimean Tatar village of Karakuyu in the province of Ankara), Turkmen (925 in the province of Tokat), Uyghur (500+ in Kayseri and Istanbul), and Uzbek (1,981 in the provinces of Hatay, Gaziantep, and Urfa).

Within the Indo-European family, the largest representatives in Turkey until the early twentieth century were Greek and Armenian, which had about two million speakers each in Turkey in 1900. At the present time, there remain a few tens of thousands of Greek speakers in Turkey (mostly Muslims from the eastern Black Sea area; Orthodox Greeks now number only 3,000 in the entire country), and less than 50,000 speakers of (Western) Armenian, almost all of them in Istanbul.

There are also dozens of villages in Artvin, the northeasternmost province of Turkey, in which varieties of Homshetsma, a western Armenian dialect, are still spoken. The exact number of Hemshinli who still speak their ancestral language is unknown; 20–30,000 would be a rough estimate. All Hemshinli speak Turkish as well, and the western Hemshinli (who live in the Rize province) now speak only Turkish, peppered with remnants of their original Armenian dialect (Bläsing 1992). This change was precipitated by the conversion of the Hemshinli from the Armenian Apostolic church to Islam beginning in the sixteenth century; those Armenians who did not wish to convert fled primarily to the Black Sea coastal regions of Abkhazia and Russia, where they still reside today. Most Hemshinli at present do not consider themselves to be Armenian, and many of those who speak Homshetsma are unaware that it is a form of Armenian, a belief that is facilitated by the large number of Turkish loanwords and the Turkic-like structure of the grammatical system.

The Boshas or Poshas, sometimes referred to as Armenian gypsies (because of their language, not their ethnicity), also reside in Turkey (chiefly Istanbul; there are also communities in Erevan, Armenia, Javaxk, Georgia, along the Black Sea, and in other areas fomerly inhabited by Armenians) and speak a language related to Armenian. This language, Lomavren (containing Lom 'gypsy' + the Armenian language suffix -(e)rēn), is a 'mixed language', wherein the lexicon is largely Indic (as with all gypsy languages) and the grammar and phonology are based on the Armenian dialect of Erzerum in northeast Turkey, where the Boshas once resided. As of 1982, there were also at least two other gypsy languages spoken natively in Turkey: approximately 20,000 speakers of Domari, mainly in western Turkey, and some 40,000 speakers of Balkan Romani (Arlija).

The Indo-European branch that is best represented in Turkey today is Iranian, thanks primarily to the several million speakers of Kurmanji and other dialects of Kurdish, but also Zaza (or Dimli; one million speakers in 1992), Herki, and Shikaki. Bilingualism with Turkish among the Iranian populations in Turkey is not as common as it is with the other minority languages, presumably due to the significantly greater size and viability of the community, geographical obscurity, and the lower literacy rate (28%) than in the population as a whole (about 70%).

Speakers of the northeast Iranian language Ossetic have been reported in Bitlis (Digor dialect) and Sarıkamış and Erzerum (Iron dialect). Speakers of various Ossetic dialects have also been reported in Muğla, Kars, and Antalya.

At least two Balkan Slavic languages are represented in Turkey as well: Bulgarian (270,000 speakers in 1993, scattered in Edirne and other western provinces) and Bosnian (20,000 first language speakers in 1980, out of an estimated 61,000 Bosnian Turks scattered through western Turkey). It is not entirely clear how these populations have been affected by the upheavals in Bosnia and Kossovo in the 1990s, but there was at least a temporary exodus to Turkey during that period; the same holds for the Albanian speakers in Turkey, who numbered about 65,000 in 1993, scattered over western Turkey.

The Anatolian branch of Indo-European was located entirely in what is now called Turkey. The family includes Hittite, Luvian, Lycian, Lydian, Palaic, and Carian, and appears to have completely died out by about the end of the first millennium BC. Extensive collections of cuneiform tablets (in Hittite and Luvian), heiroglyphic inscriptions (in Luvian), and inscriptions in Greek letters (in Lycian, Lydian, Palaic, and Carian) were unearthed and deciphered in the twentieth century. The Indo-European language Phrygian was also spoken in the western half of Turkey, and it too is now long extinct.

Two Caucasian language families are fairly well represented in Turkey, partly due to spillover at the Georgian border in northeastern Turkey, and partly thanks to the massive influx of refugees from the northwest Caucasus following the defeat of the Caucasian confederation by Russia in 1864, as a result of which the Czar deported roughly half of the Northwest Caucasian population to the Ottoman Empire. Some Northwest Caucasian languages survived only in the diaspora, such as Ubykh (now extinct) and the Sadz dialects of Abkhaz. Other Northwest Caucasian languages still to be found in Turkey today include Abaza (10,000 speakers in Turkey as of 1995), Abkhaz (35,000 speakers in Turkey as of 1993, mainly in Artvin in northeast Turkey and villages in the provinces of Bolu and Sakarya), Adyghe (130,000 in 1965, in various provinces in central and western Anatolia), and Kabardian (202,000 speakers in Turkey as of 1993, primarily around Kayseri).

Northeastern Turkey contains the westernmost component of the South Caucasian language family, notably Laz (92,000 speakers in Turkey in 1980 along the Black Sea Coast of the provinces of Rize, Artvin, Sakarya, Kocaeli, and Bolu) and Georgian (some 40,000 first language speakers in 1980 out of 91,000 ethnic Imerxev, a Georgian minority, located primarily in the provinces of Artvin, Ordu, and Sakarya).

Several Semitic languages survive in Turkey, including Arabic (400,000 speakers as of 1996), Turoyo/Suryani (3,000 speakers in 1994, originally in

the Mardin province), and Hertevin (250–300 speakers in 1994, originally in the village of Ekindüzü in the Siirt province). Turoyo is a modern descendant of Syriac, the language of the Syrian church, which was centered in Edessa (modern Şanlıurfa, in southeastern Turkey) and spoken until about the twelfth century AD, although it is still used as a literary secular language by some members of the Syrian church.

One may also classify the language of the Sephardic Jews, Ladino (or Judeo-Spanish, as it is sometimes called) as partially Semitic, although its core is an Iberian Romance language closely related to Spanish. As of 1980, it was estimated to have 8,000 or fewer first language speakers, mainly in Istanbul but also in Izmir. The Dönme, adherents of Shabbetai Zevi who converted to Islam, also speak Ladino; there were 15,000 of them residing in Turkey in 1976. Almost all speakers of Ladino are bilingual in Turkish.

Turkey also possesses at least one distinct signed language, Turkish Sign Language. To the best of my knowledge, no linguistic studies have been carried out on this language.

References

Benninghaus, Rüdiger, and P. Andrews (eds.) 1989. Ethnic groups in the Republic of Turkey. Wiesbaden: Ludwig Reichert Verlag.

Bläsing, Uwe. 1992. Armensiches Lehngut im Türkeitürkischen am Beispiel von Hemşin. Amsterdam: Rodopi.

Barak, Veysi and Joyce Blau. 1999. Manuel de Kurde: Kurmanji. Paris: L Harmattan.

Chirikba, Viacheslav. 1996. Common West Caucasian. Leiden: CNWS Publications.

Dumézil, Georges. 1963. Notes sur le parler d'un arménien musulman de Hemshin. Mémoires de l'Académie Royale de Belgique. Classe des Lettres 57.4. Brussels.

Finck, Franz Nikolaus. 1907. Die Sprache der armenischen Zigeuner. St. Petersburg.

Hadank, Karl. 1932. Mundarten der Zâzâ. Berlin: Verlag der Preussischen Akademie der Wissenschafter in Kommission bei Walter de Gruyter & Co.

Kornfilt, Jaklin. 1997. Turkish. London: Routledge.

Newton, Brian. 1972. The generative interpretation of dialect: a study of modern Greek phonology. Cambridge: Cambridge University Press.

Vaux, Bert. 1998. The phonology of Armenian. Oxford: Oxford University Press.

BERT VAUX

Typology

At its core, typology is the classification of languages on the basis of shared formal characteristics. In this way, it is distinct from genetic classification, which attempts to determine shared ancestry among languages, and from demographic classification, which categorizes languages based on geography, the numbers of speakers, or other social variables. The ultimate goals of typology are to ascertain the ways in which languages are similar in structure and to determine just how different human languages can be. Concomitant to these goals is the discovery of traits in language that are logically distinct but actually occur together.

It should be stressed that typology is not a theory of language structure. It does not attempt to provide a formal model of sentence structure or sound systems. Rather, typology is an approach that one can take in investigating the composition of human languages, an approach that is driven by a method of crosslinguistic comparison and rests on the assumption that structural similarities between languages disclose fundamental properties of human language more generally.

Universals

One of the core notions of typology is that of the 'language universal', a statement about what is always or nearly always true about language structure. Universals that are without exception, such as 'all languages possess both vowels and consonants', are called 'absolute'. Universals that admit exceptions but hold true to a statistically significant degree are referred to as 'probabilistic' universals. An example of a probabilistic universal is 'Languages nearly always distinguish between nouns and verbs'. While noun–verb polarity is a dominant characteristic in languages, there are a few known languages, such as Tongan, which lack it; hence, the property does not hold as an absolute universal.

The universals noted so far assert the presence of a single structural property, and as such they logically

define two types of languages: those that have the property and those that lack it. The latter is only possible in the case of probabilistic universals. However, many typological universals are stated in the form of 'implications'. That is, the statement of the universal carries a condition. An example of an implicational universal is: 'Languages with dominant Verb Subject Object order are always prepositional' (Greenberg 1966). Here the property of having prepositions (as opposed to postpositions) is said to depend on another property, namely having a certain word order in the sentence. This implication is absolute, but many implicational universals are probabilistic as well.

Both simple and implicational universals reveal something significant about the form language takes, but in different ways. Simple universals bring language structure into relief against ways we might imagine it could be different. We can create a mental construct of a language that lacks consonants, for instance, but such a language simply does not exist. Implicational universals, on the other hand, underscore connections between language properties that are logically distinct. There is no a priori reason that the order of subjects, objects, and verbs should be tied into a language using prepositions, yet there the correlation exists.

Sampling

For many reasons, not all human languages can be simultaneously compared. Many languages have died leaving behind no trace, and many languages have not yet come into being. Detailed descriptions are only available for a fraction of known languages. Therefore, by necessity, typologists must create a sample of languages when they carry out their research.

Since most typological universals are probabilistic, the structure of the sample is of great import. If the languages in the sample are not selected to protect against certain biases, then the information that is gleaned will be unreliable. Two main types of bias must be controlled for when constructing a sample: a genetic bias and an areal bias. Languages that are in the same language branch or language family may be similar because they have retained features from a common ancestor. If a particular family is represented in the sample disproportionately, then a pattern might emerge that reflects the properties of this language family rather than languages in general, and thus the sample would be genetically biased. An areal bias can arise in a language sample when a disproportionate number of languages from one geographical region is used. Since languages that are in close proximity tend to influence each other's structure, certain structural features come to be shared by all languages in a given area. By placing too many languages from one area

into a sample, any shared areal features will be overrepresented in the data.

Word-Order Typology

Although typological studies have probed into many facets of language structure, the most prominent focus of research in the last 50 years has been on word order. Joseph Greenberg's seminal publication in 1966, 'Some universals of grammar with particular reference to the order of meaningful elements', identified 45 universals, 28 of which dealt with the relative order of sentence elements. Greenberg's universals were of the implicational sort; e.g. he proposed, 'If the nominal object always precedes the verb, then verb forms subordinate to the main verb also precede it' (1966:110).

In Greenberg's universals, there is repeated reference to the ordering of subject, object, and verb in the precondition of the universal. This striking fact led other linguists, in particular W.P. Lehmann, to propose that the relative ordering of verbs and objects was in some sense basic to language structure, and the relative ordering of other constituents was largely predictable once the order of verb and object was known. For instance, according to Lehmann, in languages with the most common order Verb–Object, one would more than likely find prepositions before the nouns they governed, nouns before modifying adjectives, negatives before verbs, and comparative adjectives before the standards of comparison. In languages where the most common order is Object–Verb, just the opposite orderings are predicted to occur.

There have been four further developments of note in word-order studies. First, efforts have been made to recast the correlations like those explored by Lehmann in terms of more abstract grammatical concepts such as 'head' and 'modifier' instead of treating verb vis-à-vis object as the dominant organizing principle. Second, there has been much discussion regarding why such correlations between sentence elements exist and regarding the ways individual languages depart from the word-order patterns that normally hold sway. Third, in 1983 the linguist John Hawkins attempted to provide a fuller set of absolute universals of word order that would allow linguists to get a better idea of possible, typical, and impossible language types. In order to formulate his universals in absolute terms, Hawkins used multiple preconditions. For example, one of Hawkins's universals claims that if a language has Subject–Object–Verb order, AND the language puts adjectives before nouns, then the language will also put genitives (possessors) before their governing nouns. Finally, by improving on the sampling methods employed in earlier word-order studies, linguists have provided a much better basis for claims

about correlations. Notable in this area is the work of Matthew Dryer, who has demonstrated, among other things, that earlier claims about the correlation between verb/object ordering and noun/adjective are suspect, as is the correlation between verb/object and negative/verb.

Morphological Typology

Typologists of the nineteenth century wondered whether languages could be categorized on the basis of their word structure (morphology). This early work was revised by the great American linguist Edward Sapir, who proposed that languages could be classified on the basis of two independent parameters: how many word elements are permitted within individual words, and how much the pronunciation changes when word parts are put together. With respect to the first parameter, languages are said to be 'analytic' (or 'isolating'), i.e. words are not divisible into parts; 'synthetic', i.e. manifesting a limited number of word parts per word; or 'polysynthetic', i.e. employing many parts per word. With respect to the second parameter, nonisolating languages are either 'fusional', where the word parts are relatively difficult to isolate from each other and show considerable variation in their sound structure depending on the particular combinations; or 'agglutinative', meaning that they have easily identified word parts.

Because these labels appear to be impressionistic rather than categorical, and because no language is of a pure type (for example, agglutinative languages invariably have some fusional aspect), contemporary typologists have paid relatively little attention to developing Sapir's morphological typology. Even so, the terminology is widely employed to provide broad characterizations of languages in terms of their overall word structure.

In recent years, far more attention has been paid to a different parameter related to word structure, 'head marking' vs. 'dependent marking', a distinction explored most thoroughly in the work of Johanna Nichols. In constructions involving a head element and an element that depends on it, languages vary with respect to how their relationship is indicated. Consider possessive constructions, where the possessor is the dependent and the possessed element is the head. English, for example, is said to use dependent marking to express the dependency, because the possessor appears either with the possessive *-s* (*John's house*) or in the form of a prepositional phrase (*the house of John*). Other languages mark the head in this construction (e.g. Hungarian *az ember ház-a* 'the man's house', literally 'the house man-his'). Still other languages mark both the head and the dependent, or do not mark either of them. Languages can then be placed into types based on the general preference they display for head marking, dependent marking, double marking, or no marking. Once so categorized, links between this parameter and other structural traits can be explored.

Explanations for Universals

Although typology is not a theory of language per se, the discipline has traditionally been aligned with functionalist and historical approaches to language. The link is most clearly revealed in the kinds of explanations that typologists offer for the universals they have discovered. Rather than provide explanations based on the system of language itself, typological explanations are typically external to the language system. That is, the explanations rest on considerations about how language is used, how it is perceived, how it is processed, and how it is shaped by human cognition. Moreover, the explanations tend to be cast in a historical perspective. It is assumed that a language comes to have the shape it does because of the interaction of regular processes of historical change with functional pressures that nudge speakers to maximize the efficiency of their languages for communication.

To better exemplify the nature of typological explanations, consider the phenomenon of 'pro-drop'. In many languages where the verb agrees with the subject, a subject pronoun need not be included. In Spanish, for instance, the sentence *Yo bailo* 'I am dancing' is more commonly rendered *Bailo*, without the subject pronoun *Yo* 'I'. Because information about the subject is included in the verb (the ending *-o* indicates a first person singular subject), the dropping of *Yo* causes no loss of information. In contrast, English verbs show very little agreement with subjects, and the dropping of a subject pronoun is not permitted, presumably because information about the subject cannot be recovered from the verb. This link between subject agreement and the ability to drop subject pronouns is a strong one. Typologists see the link as part of a larger functional pressure on language referred to as 'economy'. Languages tend to be economical, eliminating redundant information where possible. Over time, in a language where information is being expressed redundantly in the grammar, the grammar changes to eliminate the redundancy.

The functional pressures operating on language are often expressed in terms of 'markedness'. The core idea behind markedness is that there are asymmetrical relationships between grammatical elements that are otherwise equal. For example, the sounds [p], [t], and [k] (as well as several other sounds) form a natural class in that they are all voiceless stops. As members of this set, they all have equal status. However, when one examines the set of voiceless stops that occur in the sound systems of languages, it becomes apparent

that there is something special about [t]: If a language only has only one voiceless stop, it is very likely to be [t]. And when one considers [p] and [k], one finds something special about [k]: if a language has [p], then it also has [k]. Therefore, although these sounds are equal in terms of their membership in the natural class of voiceless stops, they are unequal in the way that they are distributed in languages. [t] is then said to be 'unmarked' relative to [k] or [p]. Furthermore, [k] is unmarked relative to [p]. When a series of these markedness pairs occurs within a set of otherwise equal elements, the serial relationships can be expressed in terms of a 'markedness hierarchy'. For the case at hand, the hierarchy is $t<k<p$, with the least marked member furthest to the left on the hierarchy.

Such hierarchies are commonplace in typology. Some of the better known hierarchies include the grammatical relations hierarchy (subject < object < indirect object < other), the number hierarchy (singular<plural<dual<trial/paucal, i.e. 'three/few'), the animacy hierarchy (first and second person pronouns < third person pronouns < proper names < human nouns < animate nouns < inanimate nouns), and the color term hierarchy (white/black < red < green/yellow <blue < brown). While all these hierarchies obviously have quite different implications for what the structure of language is like, they all highlight the sometimes surprising asymmetry that arises between grammatical elements in language.

References

Comrie, Bernard. 1989. Language universals and linguistic typology, 2nd edition. Chicago: University of Chicago.

Croft, William. 1990. Typology and universals. Cambridge: Cambridge University Press.

Dryer, Matthew S. 1992. The Greenbergian Word Order Correlations, Language. 68. 81–138.

Greenberg, Joseph, H. 1966. Some universals of language with particular reference to the order of meaningful elements. Universals of grammar, ed. by Joseph H. Greenberg. Cambridge, MA: MIT Press.

Hawkins, John A. 1984. Word order universals. New York: Academic Press.

Mallinson, Graham, and Barry J. Blake. 1981. Language typology. Amsterdam: North-Holland.

Nichols, Johanna. 1986. Head-marking and dependent-marking grammar. Language. 62. 56–119.

Sapir, Edward. 1921. Language. New York: Harcourt, Brace and World.

Shopen, Timothy (ed.) 1985. Language typology and syntactic description, 3 vols. Cambridge: Cambridge University Press.

Whaley, Lindsay J. 1997. Introduction to typology: the unity and diversity of language. Thousand Oaks, CA: Sage.

LINDSAY WHALEY

See also **Greenberg, Joseph Harold; Lehmann, Winfred Philipp; Sapir, Edward**

U

Udmurt

Udmurt, also known as Votyak, is a language with an old written literature, which has long been neglected by scholars, politicians, and even the inhabitants of the Udmurt Republic itself. The alternative name was promulgated via Russian and is now considered old-fashioned; indeed, many ethnic Udmurts also consider it disparaging. The word *Udmurt* is formed by the compounding of *murt* 'human being' and **odo* (the latter is of obscure origin).

The earliest extant literary texts in Udmurt date from the sixteenth century and are typically word lists. The first grammar of Udmurt was published at the end of the eighteenth century, while in the mid-nineteenth century the Gospel texts of the New Testament were translated. Following the Great October Revolution and the inception of the Communist state, many other works of literature have been translated into Udmurt, including the leading texts of Marxism–Leninism.

The Udmurts and their language developed between the Volga-Kama and Vyatka Rivers during the second half of the first millennium of the Christian era and there has been a distinct Udmurt nation since c. 1100. In terms of genealogical classification, the Udmurt language belongs to the Permian branch of the Finno-Ugric languages of the Uralic family, and is most closely related to the Komi languages: Komi-Zuryanskiy and especially Komi-Permyalskiy. The two peoples separated between c. AD 900 and 1100 and today the two languages (Udmurt and Komi) are not mutually comprehensible; however, they occasionally present noticeable similarities in vocabulary. For example, *night* is Udmurt *uj* and Komi *voj*. The oldest Permian language, Old Permian, cannot really be said to be the precursor of any one of the modern Permian languages; however, it is considered to be closer to Komi-Zuryanskiy (also known as Komi, the language of the Komi Republic) than to Udmurt. According to scholars, the relationship between Udmurt and Komi is similar to that which holds between Russian and Polish. The phonetic system and grammatical structures of these languages are almost identical, and they have much common vocabulary. Through the centuries, the Udmurt people have come into contact with many different peoples and their languages and this has affected their language. During the first half of the second millennium of the Christian era, the Tatars were the most significant contacts; however, during recent years, the influence of ethnic Russians has proved more decisive. As a result, loanwords from Russian, or through Russian, are common: for example, *tsement* 'cement', *t'ipovoj* 'typical', or *sborn'ik* 'collection'.

Udmurt is spoken principally in the Udmurt Republic, which lies some 1,325 kilometers east of Moscow in the Pre-Ural region of the Russian Federation. Approximately, two thirds of all Udmurts live in the Republic, where they constitute c. 30% of the population. The remainder of the population of the Udmurt Republic are predominantly ethnic Russians. The final third of Udmurts live chiefly in the Perm, Kirov, and Yekaterinburg provinces of the Russian Federation and also Tatarstan and Bashkorstan. It is difficult to quote a precise figure for the total number of ethnic Udmurts; however, a population of between approximately 600,000 and 750,000 is generally agreed upon and this concurs with the 1989 Russian Federation census statistics, which indicated a poten-

UDMURT

tial population of c. 740,000 Udmurts. Where agreement ends, however, is on the question of the number of Udmurt speakers.

Udmurt had never been an official state language, and by the beginning of the twentieth century it had all but died out. It was not until the 1920s and 1930s that there was any concerted effort to preserve the language, and the culmination of this may be seen in the 1994 constitution. In this document, Udmurt is presented as one of the two official languages of the Udmurt Republic, along with Russian. However, according to some sources, there has been considerable opposition to the state sponsorship of the language, and now those areas of life where Udmurt is used and where it holds prestige are limited. There is considerable debate as to the current number of speakers of Udmurt, with figures ranging from in excess of 500,000 at one extreme to under 50,000 at the other end of the spectrum. Both sets of figures should be viewed with caution until reliable statistical data can be researched. However, UNESCO classifies Udmurt as an 'endangered' language and it would be safe to say that Udmurt is not a language enjoying a renaissance at the present time.

Udmurt is written using a modified version of the Cyrillic alphabet consisting of 38 characters. However, prior to 1930 both the Latin and the Cyrillic alphabets had been used. The language can be divided into four main dialects and these are typically referred to as the Northern, Southern, Central, and Peripheral dialects. The first three of these groups refer to the geographical location, within the Udmurt Republic, of the speakers, while the final group is the term used to refer to those speakers who live outside the Republic, notably in Tatarstan and Bashkorstan. These dialects are mutually comprehensible, although they do differ to some degree from each other in terms of vocabulary and, to a lesser degree, pronunciation.

Udmurt has seven vowels and 26 consonants. Syllables can be formed by combining consonant (C) and vowel (V) sounds in a way that can be presented schematically as follows: (C(C))V(V)(C(C))—segments in parentheses are optional.

V:	*api* (noun)	'elder sister'
CV:	*ju* (noun)	'grain'
CVC:	*tus* (noun)	'appearance'
CCV:	*visjas'kyny* (verb)	'to detach oneself'
VC:	*an* (noun)	'chin'
VCC:	*akt* (noun)	'document' (loanword from Russian)
CVV:	*tue* (adverb)	'this year'
CVCC:	*šal'kketyny* (verb)	'to flap'
CVVC:	*kuat'* (numeral)	'six'

Thus, it can be seen that syllables require a vowel sound, which may be either long or short. The core vowel sound may be preceded, succeeded, or framed by (clusters of) consonant sounds. The one proviso is that consonant clusters may not exceed two consonant sounds. Word stress typically falls on the final syllable of the word; however, in certain circumstances the first syllable carries the main stress. This is usually the case, among others, with imperatives and adverbs and adjectives, which have been created through a process of reduplication: for example, *gordgord* 'bright red'.

Udmurt has six parts of speech: adjective, noun, numeral, particle, pronoun, and verb. These categories may then be subdivided further. Nouns in Udmurt are inflected for case, number (singular and plural), and possession. Udmurt has 15 grammatical cases, which are marked via affixes. Besides indicating whether a noun represents the subject or object of a clause, the cases of Udmurt are furthermore used to encode information such as 'from…', 'to…', 'with…', etc. Possessive inflectional affixes may also be added to adjectives, nouns, numerals, and pronouns to provide information on number (i.e. singular and plural) and person (first, second, or third). Interestingly, there is no case that corresponds to the Indo-European vocative (used to mark the addressee of an utterance). Udmurt verbs fall into two conjugation patterns, depending on the last vowel of the stem. Verbs in Udmurt are inflected for mood, negation, nonfinite categories (gerund, infinitive, and participle), person, and tense.

In affirmative intransitive sentences, the typical word order is: subject–(adverb)–verb, or in short S–(A)–V. Since the copula (i.e. 'to be') is not required to be used in the present tense, this structure may be reduced to S–A. In affirmative transitive sentences, the typical unmarked word order is: subject–(adverb)–object–verb, or S–(A)–O–V.

Negation in Udmurt is typically expressed via a negative auxiliary or negative particle, which is placed before the verb; thus, S–(A)–(O)–Neg–V.

The word order in imperative sentences is usually verb initial, except in those cases where a personal pronoun is present. In the latter case, the personal pronoun is initial: (PersPro)–V. When the imperative is negated, the negative is placed in clause initial position: i.e. Neg–V.

Interrogative sentences in Udmurt appear to have no set word order pattern, since interrogation may be marked in a number of ways: an interrogative particle may be affixed to the word which forms the focus of the question; or this word may be stressed to mark the sentence as a question. However, some tendencies may be discerned. Interrogation sentences with a question word normally follow the pattern Q–S–V (where Q is the question word).

Adverbial phrases in Udmurt may appear at the beginning, middle, or end of a sentence. Adverbial positioning may be to an extent determined by the adverbial category; thus, adverbs of time and place

often occur in sentence-initial position, while certain other adverbs may appear immediately before the object or the verb.

References

Alatyrev, V.I. (ed.) 1970. Grammatika sovremennogo udmurtskogo yazyka, Sintaksis prostogo predlozheniya. Izhevsk: Izdatelstvo Udmurtiya.

Campbell, George L. 1991. Compendium of the world's languages: Vol. 2, Maasai to Zuni. London: Routledge.

Comrie, Bernard. 1981. The languages of the Soviet Union. Cambridge: Cambridge University Press.

Government of Udmurtia. 1995. Konstitutsiya Udmurtskoj Respubliki. Izhevsk: Udmurtia Press.

Katzner, Kenneth. 1995. The languages of the world, new edition. London: Routledge.

Lallukka, Seppo. 1990. The East Finnic minorities in the Soviet Union. Suomalainen tiedeakatemia: Helsinki.

Suihkonen, Pirkko. 1990. Korpustutkimus kielitypologiassa sovellettuna udmurtiin (Computer corpus analysis in language typology applied to Udmurt), Suomalais–Ugrilaisen Seuran Toimituksia 207. Helsinki: Suomalais–Ugrilainen Seura.

———. 1995. Udmurt–English–Finnish dictionary with a basic grammar of Udmurt (Lexica Societas Fenno-Ugricae XXIV). Helsinki: Suomalais - Ugrilainen Seura.

UNPO. 2002. Udmurt (www.unpo.org/member/udmurt/ udmurt. html).

Vaba, Lembit, and Juri Viikberg. Udmurts or Votyaks, Endangered Uralic peoples: short reference guide, ed. by Andres Heipanuu, Information Centre of Finno-Ugric peoples: Tallinn (http://www.suri.ee/eup/udmurts.html).

Winkler, Eberhard. 2001. Udmurt (Languages of the world/ materials 212). Munich: Lincom.

KARL BERNHARDT

United States

The United States is a country of some 250 million people spanning a large portion of the North American continent. Language has been a controversial political subject in the United States since the time of Noah Webster, and the controversy continues in debates about language in the schools and the unofficial status of English today. Although today it is an essentially monolingual English-speaking country, the United States has a rich and varied linguistic heritage which includes indigenous languages, immigrant languages, American Sign Language, and a wide variety of English dialects.

Indigenous Languages

At the time when settlers of European descent arrived in the territory that is now the United States, there were at least 300 languages spoken there. Most of these languages are either extinct or endangered today, with fewer than 50 having more than 1,000 speakers. Those with more than 10,000 speakers (data from 1977–1990) are listed below:

Indigenous Languages with More Than 10,000 Speakers

Language	Number of Speakers	Family
Papago-Pima	12,000	Uto-Aztecan
Western Apache	13,000	Na-Dené
Central Yupik	15,000	Eskimo-Aleut
Dakota	15,000	Siouan
Cherokee	22,500	Iroquoian
Western Cree	35,000	Algonquian
Western Ojibwa	35,000	Algonquian

Language families with the widest geographical distribution in the United States include Algonquian, Aztec-Tanoan, Eskimo-Aleut, Macro-Siouan, Na-Dené, and Penutian. Hawaiian, an Austronesian language, is spoken by fewer than 1,000 people today.

The serious study of aboriginal languages of the Americas, following the lead of Franz Boas, in combination with developments stemming from European philology, formed the basis for modern linguistics. The following statement is from the resolution on Language Rights passed by the Linguistic Society of America.

> Many past and present members of the Society have devoted their professional lives to documenting and analyzing the native languages of the United States. Unfortunately, most of the indigenous languages of the United States are severely threatened. All too often their eradication was deliberate government policy…. The decline of America's indigenous languages has been closely linked to the loss of much of the culture of their speakers. Because of this history, the Society believes that the government and people of the United States have a special obligation to enable indigenous peoples to retain their languages and cultures. The Society strongly supports the federal recognition of this obligation, as expressed in the Native American Languages Act.

This Act of Congress states that 'It is the policy of the United States to… preserve, protect, and promote the rights and freedom of Native Americans to use, practice, and develop Native American languages'. This is a major reversal in policy from the days when Indian children were removed from their tribes and sent to English-only boarding schools where they were

punished for speaking their native languages. These earlier educational policies led to the loss of languages of tribes that had survived the warfare and epidemics that had destroyed so many other tribal languages. There are few monolingual speakers of indigenous languages in the United States today, a fact that suggests their impending replacement by English. Although some Native American languages are now being taught in tribal schools, it remains to be seen whether these languages can survive as a means of everyday communication.

History of American English

At the end of the eighteenth century, in the midst of anti-British sentiment born of revolutionary fervor, Americans declared verbal independence and the language that was formerly just 'English' became British English and American English. This was the first of many such splits into national varieties that were yet to come as a result of the far-reaching colonialism of the British Empire. Despite the fact that the English believed the speech of the American settlers to be corrupt and full of 'barbarisms', Webster defended it in his *Dissertations on the English language* by declaring that 'the people of America… speak the most pure English now known in the world'. His spelling reforms and insistence on middle class rather than aristocratic norms were intended as a contribution to the formation of an American identity.

In colonial times, however, a number of European languages were spoken in North America. English came to predominate, but not without feeling the influence of these, and some of the indigenous languages, upon its vocabulary. Even before the arrival of the Mayflower, words were borrowed into English from the indigenous people of the Americas, especially the Caribbean, via Spanish; other Native American words were introduced into English by French-speaking explorers. The Algonquian languages were the source of most borrowings from Native American languages, as this language family included the speech of the tribes along the Eastern Seaboard that were the first ones encountered by the English settlers. The settlers coexisted fairly peaceably with the Indians for the first few years, and they had need of words to describe native flora and fauna that were not found in the Old World. As hostilities worsened and settlement moved westward, most borrowings were limited to place names.

Early on, words entered English from contact with the Dutch in New York and the Germans in Pennsylvania. More lexical and even grammatical influences were to come from the many immigrant languages that arrived in later years. The impact of these immigrant languages helped to create some of the differences between regional dialects of English,

e.g. Yiddish in New York City, Scandinavian languages in the Upper Midwest, and, of course, Spanish in the Southwest, although the latter predates the founding of the United States and cannot be considered an immigrant language.

Main Sources of Borrowings in American English

Languages/ Language Groups	Predominant Region/Era	Examples
Algonquian	Eastern New England, 17th century	Moose, raccoon, squash, pecan
West African	South, 18th century	Yam, okra, banjo, jazz, tote
Dutch	Hudson Valley, 18th century	Cookie, coleslaw, boss, sleigh
German	Pennsylvania, 19th century	Cluck, dunk, hamburger, pretzel
French	Great Lakes, Louisiana, 18th, 19th centuries	Prairie, portage, armoire, jambalaya
Spanish	Southwest, 19th century	Coyote, marijuana, ranch, cafeteria
Yiddish	New York City, 20th century	Schmuck, kibitzer, phooey, schmooze

When the United States took over, first, the Louisiana Territory from the French and, later, Texas, California, and the Southwest from the Spanish, the nation thereby incorporated the many speakers of languages other than English who were already living in these areas. The US government allowed French and Spanish use to continue for a number of years in the local governments of states like Louisiana and California, although in at least two instances (New Mexico and Hawaii), the preponderance of non-English speakers in annexed territories became an obstacle to acceptance into statehood. During the many debates over the years about the status of languages other than English (and their speakers), critics have often overlooked the fact that many non-English speakers did not come to the United States voluntarily. In addition to the cases of French, Spanish, and Hawaiian and other indigenous languages noted above, many Chinese speakers were brought to the West Coast as forced laborers. The largest group of non-English speakers in US history also came involuntarily, brought as slaves from West Africa.

The debate among linguists over the history of African American Vernacular English (AAVE) has still not been resolved, although it has been one of the most productive areas of linguistic study in the United States for over 30 years, spurring a vast amount of research on Creole languages and on nonstandard varieties of English in the United States. The sides have been characterized as 'Anglicist' vs. 'Creolist'

hypotheses, and the most simplistic description of these points of view is that AAVE originated either from the nonstandard British-based varieties of English learned by the slaves or from a widespread Creole language that once existed in the slave-owning areas of the South. Most linguists today agree that a multiplicity of factors shaped the language of African Americans, including African substrate languages like Wolof and Yoruba, English-based Caribbean Creoles, the British dialects of white indentured servants, and developing American English dialects, both standard and nonstandard, as well as language-learning features that may be attributable to Universal Grammar. It is likely that the first slaves to be brought to the United States, who worked as domestic servants or on small farms where there was ample exposure to native speakers of English, were able to learn English just as other immigrants did. As more and more slaves were brought in, both directly from Africa and via the Caribbean, opportunities to learn from native speakers diminished as the ratio of blacks to whites grew. It is likely that a true Creole only developed along the coast of South Carolina and Georgia, where large, labor-intensive rice plantations produced demographics similar to those on the sugar plantations of the Caribbean. (This now mostly decreolized variety is known as Gullah or Geechee, and it is the source of the Afro-Seminole Creole in Texas.) Elsewhere in the South, a variety developed among African Americans that was very similar to white varieties of English in the region, due to mutual influence, and which had a Creole-like tense and aspect system (see below). This variety was taken northward to cities such as Detroit, Chicago, New York, and Philadelphia when many blacks left the South during the first decades of the twentieth century, fleeing poverty and racially motivated violence.

Features of AAVE

Copula absence
She pretty. They in the house.
Invariant/nonfinite be (habitual aspect)
He be in school every day.
Stressed been (remote tense)
He BEEN ate it. She BEEN gone.
Perfect done (completive aspect)
They done left the house.
Lack of verbal morphology
She always go to that store. The building burn last week.

While AAVE thus spread from south to north, other dialects spread with the population from east to west. Quite a lot of dialect mixing occurred in the early settlement period, as colonists in any one town might have come from various parts of England, Ireland, Wales, Germany, or elsewhere. There was probably a period of 'koineization', when a leveling process

occurred, although some scholars believe that even in colonial times, regional dialect differences existed. It is possible to distinguish four streams of settlement from different parts of the British Isles that may have played a role in the formation of the major dialects of American English. The eastern part of England was the origin of many New England settlers, while Pennsylvania and Delaware were settled by people from northern England, and the South by colonists predominantly from the southwest of England. Inland areas, including Appalachia, were settled slightly later by the people known as the Scots-Irish, or Ulster Scots, who were of Scottish descent but who came to the United States after a sojourn in northern Ireland.

Four major dialect areas of the eastern United States were defined by Hans Kurath based on data collected for the Linguistic Atlas of the United States and Canada, an ambitious undertaking begun in the 1920s that was designed to map the dialect areas of English-speaking North America. Two influential publications based on the atlas data (a 'word geography' and a volume describing the pronunciation of influential citizens in various regions) provided the basic outline of dialect areas in the United States for many years. Kurath proposed two primary boundaries and a secondary boundary as shown in the map (Figure 1), forming the North, North Midland, South Midland, and South dialect areas. Today, many linguists believe the primary and secondary boundaries to be reversed, i.e. that the major division is North vs. South, with two main subregions within each, corresponding to essentially the same areas Kurath identified but now known as Upper and Lower North and Upper and Lower South (Figure 2). The dialect areas remained fairly distinct as settlement proceeded westward as far as the Mississippi River, but west of there, their features were mixed together. Although there are some identifiable dialect regions in the West, as the most recently settled area of the United States, dialect differentiation is not as sharp as in the East. Likewise, dialect differences in the United States in general are more subtle than those in the long-populated countryside of Europe and England.

The vocabulary of American English is expanding at a tremendous rate, coinciding with the twentieth century developments of universal public education, improvements in transportation, mass media, and a consumer culture. Words for archaic features of the culture, such as premechanized agriculture, are disappearing, along with the many lexical items that simply drop out of fashion over the years. On the other hand, a great deal of vocabulary has been added through technological change, especially in the field of computers. Cultural changes have also given us brand names that are widely known and used. New words are added from many sources. Slang is spread through youth culture,
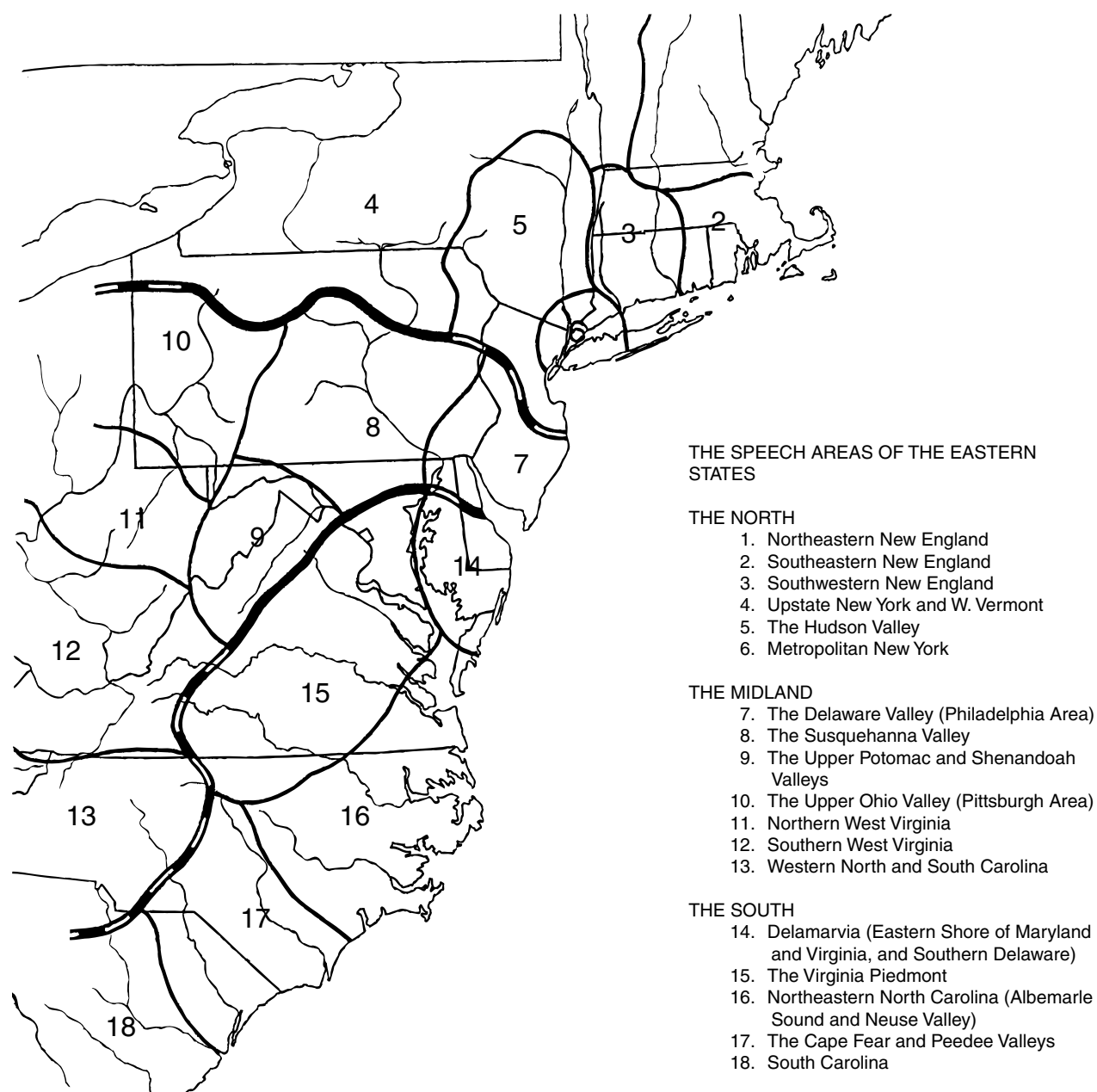
Figure 1. Major dialect areas of the US: Kurath (1949).

THE SPEECH AREAS OF THE EASTERN STATES

THE NORTH
1. Northeastern New England
2. Southeastern New England
3. Southwestern New England
4. Upstate New York and W. Vermont
5. The Hudson Valley
6. Metropolitan New York

THE MIDLAND
7. The Delaware Valley (Philadelphia Area)
8. The Susquehanna Valley
9. The Upper Potomac and Shenandoah Valleys
10. The Upper Ohio Valley (Pittsburgh Area)
11. Northern West Virginia
12. Southern West Virginia
13. Western North and South Carolina

THE SOUTH
14. Delamarvia (Eastern Shore of Maryland and Virginia, and Southern Delaware)
15. The Virginia Piedmont
16. Northeastern North Carolina (Albemarle Sound and Neuse Valley)
17. The Cape Fear and Peedee Valleys
18. South Carolina

and through mass-distributed music and other electronic media; some slang words are ephemeral, but others, such as *cool* and *dude*, have persisted for many generations. And borrowings from immigrant languages continue, although these terms may be regionally limited. Despite increasing geographical mobility and exposure to other dialects, regionalisms in the vocabulary do remain, as documented in one of the most significant American dictionary projects, the multivolume *Dictionary of American regional english*. Another

important recent lexicographical work on American English is the *Historical dictionary of American slang*. These two projects expand our resources on American English far beyond the reach of the best current unabridged dictionaries with an American perspective, *Webster's third* and *Random House*.

American English grammar has diverged from that of British English in a number of ways, for example, in retaining many uses of the subjunctive and in the use of auxiliary verbs.
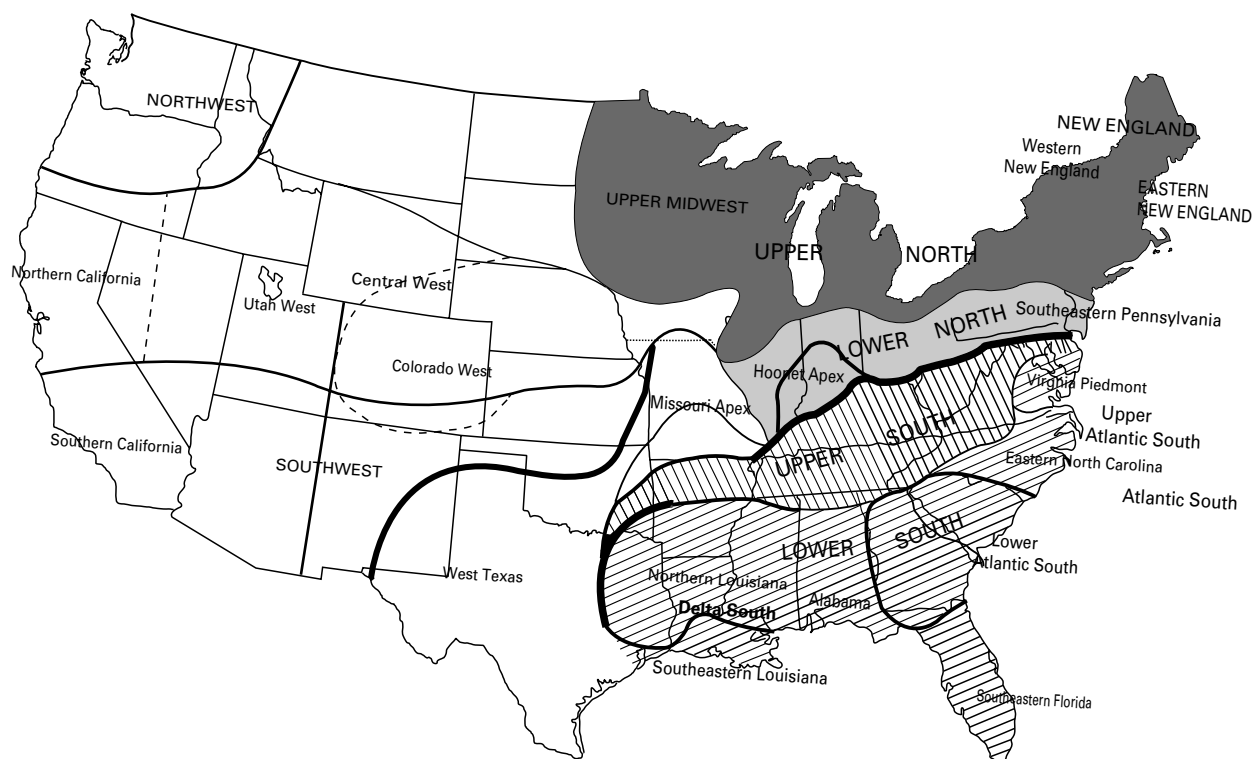
Figure 2. Major dialect areas of the US: Carver (1987).

Some Grammatical Differences between US English and English English

US English	English English
The salad has olives in it.	The salad has olives in.
In the future, we will change the plan.	In future, we will change the plan.
John gave it to me.	John gave me it.
Do you have any apples?	Have you any apples?
I won't be able to go.	I shan't be able to go.
If I were you, I wouldn't do it.	If I was you, I wouldn't do it.

Standard English grammar is that which is used by the most educated, upper social classes in all regions and which carries the prestige of being regarded as 'correct' or 'good' English by all. English syntax and morphology vary in nonstandard dialects across the United States, with much of the variation involving verb forms in the past and past participle. Nonstandard speech is usually stigmatized because it is typical of the working class or minority ethnic groups. These nonstandard dialects of English in all parts of the United States exhibit many of the same features, including lack of subject–verb agreement, as in *he don't*, *we wasn't*, and multiple negation (*they don't know nothing*), often including *ain't*.

Pronunciation changes continue to spread across the United States, although dialectologists have moved from a 'wave' model of understanding change as moving outward in an even pattern from a geographic origin (like waves in a pond) to a model where change moves from one large urban area to another before diffusing into suburbs, smaller cities, and finally rural areas. One important survey conducted by William Labov and colleagues, the Phonological Atlas of the United States, has documented several sound changes in the form of vowel shifts and mergers. In most of the United States, the vowels in *lot* and *law* are merging, and other shifts are taking place in some Northern cities (e.g. the raising of the vowel in *bag* toward the vowel of *beg*) and in the South (e.g. the merger of the vowels of *pen* and *pin* before nasal consonants) (Figures 3 and 4).

There is no single standard of pronunciation in the United States, but rather a number of 'regional standards', characterized by the accents of the upper-middle class. This includes people like lawyers, doctors, bankers, and politicians, economically successful and highly visible members of the community who often speak with a regionally identifiable accent that is the local prestige norm. Although people who move from one region to another may drop features of their native dialects that are stigmatized in other regions, and broadcasters are trained to do the same thing, this accentless type of speech is not required for success even at the national level: witness the speech of recent

/i/ (*beet*)   /u/ (*boot*)

  /ɪ/ (*bit*)   /ʊ/ (*put*)
     →
  /e/ (*bail*)   /o / (*boat*)

      /ɛ/ (*bet*)   /ʌ/ (*but*) →

      /æ/ (*bat*)   /ɔ/ (*brought*)
         →
           /a/ (*father*)
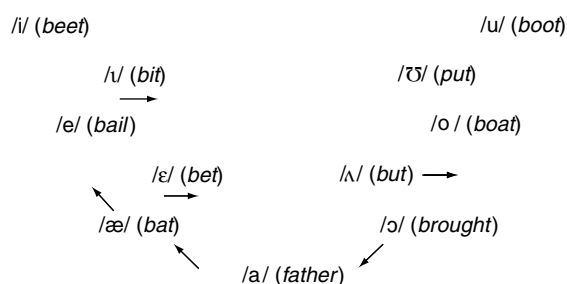
Figure 3. The northern cities vowel shift (adapted from Labov 1991).

  /i/ (*beet*)   ← /u/ (*boot*)

      /ɪ/ (*bit*)   /ʊ/ (*put*)

  /e/ (*bail*)   ← /o / (*boat*)

      /ɛ/ (*bet*)   /ʌ/ (*but*)

  /æ/ (*bat*)   /ɔ/ (*brought*)

           /a/ (*father*)
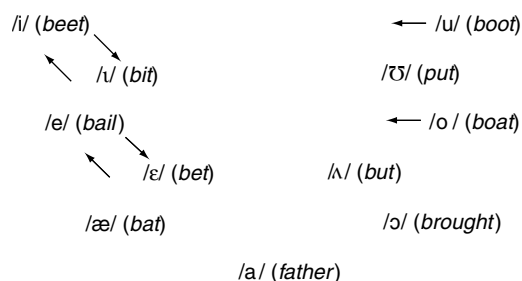
Figure 4. The southern vowel shift (adapted from Labov 1991).

presidents, the strong New England accent of the elder Bush and the Texas sound of the younger one, the Southern accents of Clinton and Carter, and the Boston accent of the Kennedys.

The speech patterns of traditionally economically depressed regions of the United States (e.g. the South), or areas with larger concentrations of working class speakers (e.g. New Jersey, the Appalachians), like the speech patterns of poor people in general, have been stigmatized on a national level. The most prestigious US dialect is one that dialectologists have labeled Inland North, which encompasses upstate New York; the northern parts of Ohio, Indiana, and Illinois; Michigan; and Wisconsin. Studies of language attitudes have shown the speech of the South to be consistently labeled as the 'least correct', although not necessarily the 'least pleasant'. Sociolinguists call this type of positive assessment of speakers of nonstandard dialects, as being more friendly and sincere, 'covert prestige'. The idea that nonstandard speech can be valued as sounding more masculine has led researchers to invoke covert prestige to explain the tendency of men in the United States to use more nonstandard features than women of the same social class. Other factors could explain this finding as well, especially the higher educational levels and Standard English skills expected of women in many low-paying jobs.

Correlational studies of how nonlinguistic variables such as social class, sex, age, region, and race influence speech patterns have taught us much about the complexity of American English. This subfield of linguistics is known as language variation study, and it includes what has traditionally been called dialectology as well as much of the subject matter of sociolinguistics. Sociolinguistics also includes the study of language attitudes, sometimes called perceptual dialectology. In many ways, the attitudes toward low-status varieties of English, especially ethnically-based varieties like AAVE, parallel attitudes toward languages other than English in the United States. These attitudes have led to political controversy over the use of languages and varieties other than Standard English, especially in the schools.

Immigrant Languages and Language in the Schools

Although the United States has no official language, English serves as the de facto national language, with 87% of the adult population speaking *only* English, and 80% of those who speak another language at home also speaking English 'well' or 'very well'. Bilingualism in the United States has mostly been 'subtractive', leading to loss of the language other than English within three generations for immigrants. The rate at which immigrants are learning English has not changed, but public perception today is that Spanish-speaking immigrants are not learning English. This is due to continued high numbers of new immigrants from Mexico and elsewhere in Latin America. Early in the twentieth century, there were complaints that Italian, Jewish, Greek, and Slavic immigrants were not learning English; the ethnic languages spoken by these groups are disappearing quickly less than 100 years later. Demographic projections predicting that Caucasians will comprise less than 50% of the population by the mid-twenty-first century have fueled fears that English speakers will become a minority in the United States. Opposition to the use of languages other than English in the public sphere by anyone besides tourists has often been linked historically to xenophobia and nativism, as when racist, anti-Catholic, and anti-immigrant political groups in the late nineteenth century fought to enact language restrictions.

Germans, who were part of the earliest European settlements in the United States, were probably more successful at maintaining their language than any other non-English-speaking group. (However, the claim that a proposed law making German the official language of the United States failed by just one vote is a myth: rather, the Congress declined, in a close vote, to take up a proposal to print laws in German as well as in English.) Although Benjamin Franklin and others complained about the use of German in Pennsylvania, German communities there and later in the Midwest continued to use German as the language of instruc-

tion in both public and parochial schools up until World War I, when nationalistic sentiment put an end to the teaching of German even as a foreign language. In some places, German textbooks were burned by mobs. Many Americans came to view bilingualism as unpatriotic and deviant, and some states even forbade the teaching of any foreign language for a time. Beliefs that bilingualism is somehow harmful persist, so that the Linguistic Society of America found it necessary to include the following in a press release on language rights: 'There is no convincing evidence that bilingualism by itself impedes cognitive or educational development. On the contrary, there is evidence that it may actually enhance certain types of intelligence.' The only places where a form of German remains viable today are the Amish communities, principally in southeastern Pennsylvania. There are, however, two other European languages that have been spoken in the United States for centuries: French varieties in Louisiana and Maine, and Spanish, which has spread from the Southwest into many urban areas.

Most Spanish speakers in the United States today are Mexican-Americans, or Chicanos. The linguistic competence of Chicanos may include a plethora of languages and varieties, as listed by Anzaldúa:

- Standard English
- Working-class and slang English
- Standard Spanish
- Standard Mexican Spanish
- North Mexican Spanish dialect
- Chicano Spanish (Texas, New Mexico, Arizona, and California have regional variations)
- Tex-Mex [a code-switched variety]
- Pachuco [an argot]

In addition, there are at least as many varieties of Spanish in the United States as there are countries of origin for Hispanics, including the prominent Caribbean dialects of Cuba and the US-governed colony of Puerto Rico. The first European settlements in what is now the United States were the Spanish settlements in Florida and Georgia. Spaniards also established missions in California very early; however, the Spanish variety that has been continued from colonial times is that which was brought into the Southwest, especially Arizona, New Mexico, and Colorado by settlers from 'old' Mexico before those states were part of the United States. Ironically, due to demographic changes and the economic power of Anglo retirees, Arizona has been one of the most active states in pushing for 'English Only' restrictions.

The English Language Amendment is a bill that has been introduced in Congress in several forms but that has never passed. It would make English the official language of the United States. There are many differ-

ent interpretations of what this would mean, but most of them involve curtailing government services or documents in languages other than English. In response to this and similar initiatives that have been passed in a growing number of states, the Linguistic Society of America passed the following resolution, mentioned above and excerpted below:

All residents of the United States should be guaranteed the following linguistic rights:

- To be allowed to express themselves, publicly or privately, in the language of their choice.
- To maintain their native language and, should they so desire, to pass it on to their children.
- When their facility in English is inadequate, to be provided a qualified interpreter in any proceeding in which the government endeavors to deprive them of life, liberty or property. Moreover, where there is a substantial linguistic minority in a community, interpretation ought to be provided by courts and other state agencies in any matter that significantly affects the public.
- To have their children educated in a manner that affirmatively acknowledges their native language abilities as well as ensures their acquisition of English. Children can learn only when they understand their teachers. As a consequence, some use of children's native language in the classroom is often desirable if they are to be educated successfully.
- To conduct business in the language of their choice.
- To use their preferred language for private conversations in the workplace.
- To have the opportunity to learn to speak, read, and write English.

This statement on linguistic rights refers to the benefits of at least some instruction in languages other than English, or bilingual education. Bilingual education is not common in the United States and where it does exist, it is usually 'transitional', meant to help students keep up with their other subjects, which they study in their native language, while they study English intensively in order to join a mainstream classroom as soon as possible. Linguists would like to see bilingual education also used for the purposes of language maintenance in order to foster a plurality of languages and cultures in the United States, especially where endangered languages are spoken, e.g. indigenous languages or Cajun French.

In Louisiana, there is a state-sponsored program, the Council for the Development of French in Louisiana (CODOFIL), which is trying to raise the status of French there and to keep it from dying out. This program sponsors activities ranging from radio

and television broadcasts in Cajun French to recruiting Francophone teachers from around the world to offer second-language classes, including immersion programs, in French. Unfortunately, it has been difficult to bridge the gap between the standard Metropolitan French of these teachers and the vernacular variety spoken by the descendants of the Acadians (who came to the United States after being persecuted and forced out of Canada). About 10% of French speakers in Louisiana speak Creole French rather than Cajun French. The CODOFIL mission statement sums up the political issues surrounding the use of minority languages in the United States:

> While we would be the last to argue that a U.S. citizen or resident does not need to speak English in order to survive and thrive in our society, we should not predicate fluency in English upon the elimination of other languages. We in Louisiana have seen first hand the effects of this linguistic genocide. Although our parents and grandparents were beaten, punished and humiliated for speaking French on the schoolgrounds, we have come to realize that full participation in American democracy does not necessarily mean forgetting our roots.

For many minority groups in the United States, there is a tension between the need to assimilate to the majority culture to attain economic and professional success and the need to maintain their ethnic identity. This is the issue that educators have tried to address by proposing that schools allow both bilingualism and bidialectalism in classrooms (as occurs, for example, in Hawaii, where half of the population speaks Hawaiian Creole; students read and write in English, but the teachers explain things in Creole). For years, schools have taken the subtractive approach to teaching Standard English, both to students whose native language is not English and to students who speak a nonstandard variety of English. This 'replacive' view assumes that Standard English is inherently superior to all other varieties, which are given labels ranging from 'substandard' and 'bad' to 'gibberish'. The stigmatization of other varieties often evokes two types of responses from students: a defensive, resistant stance toward a teacher who is trying to take away the language of their family, neighborhood, and church; or the development of low self-esteem and the internalization of the belief that their native language (and by extension, culture) is inferior. By fostering bidialectalism and bilingualism, encouraging students to be fluent in both Standard English and their vernacular, and teaching them which is appropriate in which situations, schools can lower failure and dropout rates that are tied to language difficulties.

The Native American Language Act states, 'There is convincing evidence that student achievement and performance, community and school pride, and education-al opportunity is clearly and directly tied to respect for, and support of, the first language of the child or student.' This same rationale for an 'additive' approach in the classroom can be applied to any minority language in the United States. African American Vernacular English, in particular, has been at the center of heated debates over the role of languages other than Standard English in the classroom. In 1972, linguists were involved in a court case in Ann Arbor, Michigan, where the judge ruled that AAVE was sufficiently different from Standard English to merit special treatment. Teachers were to give AAVE speakers additional help in acquiring Standard English, partly by learning about the syntactic and phonological rules of AAVE themselves so that they could point out the systematic differences to their students. A similar situation arose in Oakland, California, in 1996, when the school board adopted a policy designed to help speakers of 'Ebonics' (AAVE) learn Standard English by educating teachers about dialect differences. The public interpreted both of these attempts as a requirement to force teachers and non-AAVE-speaking students to use AAVE in the classroom (not the case), and as a plan intended to keep African American students from learning Standard English and thus denying them the social benefits that come with a fluent command of Standard English (also not the case).

Similar issues revolve around the question of the role of American Sign Language (ASL) for Deaf students. ASL was developed by educators Thomas Gallaudet and Laurent Clerc, who in the early nineteenth century introduced signs from a visual–spatial language used in France, combining them with signs already in use in the United States. ASL is not simply a signed form of English. In fact, it is grammatically unrelated to English; British Sign Language and ASL are not mutally intelligible, ASL being more closely related to French Sign Language. Linguists have described ASL in terms of its phonology (handshapes), morphology (location of sign in relation to the body, movement of articulators, and facial expressions), and syntax (sequence of signs). Children in an ASL home environment acquire ASL as their native language in the same way as other children acquire spoken languages, e.g. 'babbling' with their fingers at the appropriate stage of development. These children must become bilingual if they are to learn to read and write in English. Some members of the deaf community feel that it is an unnecesary burden on children if they are also required to learn an oral approach, which involves lip reading and the production of speech sounds. It is difficult for a deaf person to learn to pronounce a spoken language well enough to be understood by people who are not immediate family. The same issues of opportunity, identity, and community

that affect other speakers of minority languages in the United States confront bilingual ASL/English speakers, both negatively in terms of language prejudice and positively in terms of having a wide range of communicative abilities.

Languages with more than 100,000 speakers in the United States, from 1990 Census and Ethnologue.

Spanish	17,339,172	Arabic	355,150
French	1,702,176	Hindi (Urdu)	331,484
German	1,547,099	Russian	241,798
Italian	1,308,648	Yiddish	213,064
Chinese	1,249,213	Thai (Laotian)	206,266
Tagalog	843,251	Persian	201,865
Polish	723,483	French Creole	187,658
Romani	650,000	Armenian	149,694
Hawaiian Creole	600,000	Navaho	148,530
Korean	626,478	Hungarian	147,902
Vietnamese	507,069	Hebrew	144,292
Portuguese	429,860	Dutch	142,684
Japanese	427,657	Mon-Khmer (Cambodian)	127,441
Greek	388,260	Gujarathi	102,418

The economic power and cultural influence of the United States is furthering the rapid spread of English across the globe, with some 700 million fluent speakers around the world, and English functioning as an important language in 75 different countries. Many countries that formerly looked to a British model for the teaching of English as a foreign language are now adopting American English norms. Thus, it is true that the dominant language of the United States has an impact far beyond its borders. The importance of American English is not likely to decline soon, either within or outside the United States.

References

American Dialect Society website, http://www.americandialect.org

Anzaldúa, Gloria. 1987. Borderlands/La Frontera. San Francisco: Aunt Lute.

Baron, Dennis. 1990. The English-only question. New Haven, CT: Yale University Press.

Boas, Franz. 1911. Handbook of American Indian languages. Washington: Smithsonian.

Carver, Craig M. 1987. American regional dialects. Ann Arbor: University of Michigan Press.

Cassidy, Frederic G., and Joan Houston Hall (eds.) 1985-96. Dictionary of American regional English, vols. 1–3. Cambridge, MA: Harvard University Press.

Council for the Development of French in Louisiana website, http://www.codofil.org

Crawford, James. 1999. Bilingual education: history, politics, theory, and practice, 4th edition revised. Los Angeles: Bilingual Education Services.

Ethnologue website, Languages of the world, 13th internet edition, http://www.sil.org/ethnologue/ethnologue.html

Kurath, Hans. 1949. A word geography of the Eastern United States. Ann Arbor: University of Michigan Press.

Kurath, Hans, and Raven I. McDavid. 1961. The pronunciation of English in the Atlantic States. Ann Arbor: University of Michigan Press.

Lighter, J.E. 1994-97. Random House historical dictionary of American slang, vols. 1–2. New York: Random House.

Linguistic Atlas Projects website, http://us.english.uga.edu

Labov, William. 1994 and 2001. Principles of linguistic change, vols. 1 and 2 (Internal factors, social factors). Oxford: Blackwell.

Linguistic Society of America website, http://www.lsadc.org

Linn, Michael D. (ed.) 1998. Handbook of dialects and language variation, 2nd edition. San Diego: Academic Press.

Mencken, H.L. 1992. The American language, abridged edition. New York: Knopf.

Mufwene, Salikoko S., John R. Rickford, Guy Bailey, and John Baugh (eds.) 1998. African American English: structure, history, and use. London: Routledge.

Preston, Dennis R. (ed.) 1993. American dialect research. Amsterdam: John Benjamins.

Pyles, Thomas. 1952. Words and ways of American English. New York: Random House.

Trudgill, Peter, and Jean Hannah. 1994. International English, 3rd edition. London: Arnold.

Valdman, Albert (ed.) 1997. French and Creole in Louisiana. New York: Plenum.

Valli, Clayton, and Ceil Lucas. 2001. The linguistics of American sign language, 3rd edition. Washington: Gallaudet University Press,

Wolfram, Walt, and Natalie Schilling-Estes. 1998. American English. Oxford: Blackwell.

ELLEN JOHNSON

Universal Grammar

Universal grammar (UG) is a theory both about the essential nature of human language and about how language is acquired by children. The concept of UG was first suggested by Noam Chomsky in his early work on generative grammar, and it continued to evolve throughout his later writings on syntactic theory. A detailed exposition of UG appears in his 1981 work *Lectures on government and binding*. This book serves as the basis

for a particular version of UG envisioned as a system of universal principles interacting with parameters of variation. This version of UG is central to the syntactic theory, known as Principles and Parameters Theory, and has been the basis for much current research in the fields of syntax and language acquisition.

The history of UG is intimately related to conflicting theories about the nature of language acquisition. In response to the behaviorist view of language acquisition as imitation and stimulus–response learning, Chomsky argued that language acquisition did not involve learning, but was instead the result of the activation of a totally innate capacity for language, namely, UG.

Support for UG comes from observations of how children's language acquisition differs from other kinds of learning and how it is not merely imitation. First, all children acquire language at a time in their life when they cannot learn any other skill of similar complexity. Furthermore, all children achieve approximately the same degree of skill (i.e. native speaker fluency) regardless of the infinite variations in their environmental stimuli. The language children produce is not merely an imitation of the adult speech they hear around them: they often use innovative structures that they could not possibly have heard from an adult. Children are not actively taught by adults which language structures are grammatical and which ones are not; they acquire a sense of the difference merely through exposure, and they generally do not respond to corrections when they do receive them. Finally, only humans seem to be able to acquire language. Primates cannot acquire most grammatical features of language, even through intensive behavioral training.

Chomsky concluded from observations such as these that language must in some sense be an inborn ability universally present in humans. At first, he described the innate ability to acquire language by using the metaphor of a 'black box', or device in the brain: the language acquisition device (LAD). Today, the concept of UG has replaced the metaphor of the device and the term LAD. It is UG that is envisioned as a biological 'language organ', and it is the nature of UG itself that allows humans to acquire language quickly and flawlessly.

UG contains the basics that all human languages have in common. Specifically, this includes the modular organization of the language faculty itself, a set of universal principles governing the structures of human language, and a set of parameters defining the possible limits of language-specific variations. This is the picture of UG that has arisen within Principles and Parameters Theory. It provides an account for both the fundamental similarities and the vast amount of observable variation among existing human languages. It also embodies predictions about what constitutes a possible human language.

Language may be structured as a system of interacting modules or subsystems. These modules are specialized in their function, much like bodily organs, and are usually assumed to include systems devoted to vocabulary, grammar, pronunciation, and meaning. UG is not seen as a list of universally valid rules for the correct formation of language structures. As Chomsky points out, lists of individual rules cannot adequately model the nature of language competency in fluent speakers or how this competency is acquired. On the contrary, learners do not need to acquire any organizational principle for the complexities of language: the framework is already provided for them by UG, and it is the same for all languages.

UG is assumed to contain a set of universal principles that are valid for all human languages. These are thought to include principles about the way language is structured into phrases. For example, a basic template governing the shape of syntactic structures and a general rule relating different syntactic structures to each other are both assumed to be principles of UG. The principles are innate and immediately available to the learner, and they are perhaps the most crucial tool in UG in its role as a language acquisition device. Because they radically reduce the number of possible analyses that a learner might assign to any one sentence, the UG principles, in a sense, allow the learner to automatically recognize the correct structure of language input during the acquisition process.

Of course, individual human languages show a great deal of surface variation in terms of the sounds they use, the structure of words, and the order in which words are combined. The language learner's task is to acquire the specific characteristics of the language of his or her community. UG therefore is said to include a set of parameters listing which aspects of language may show variation and defining what the boundaries for such variation may be. For example, one area in which languages exhibit systematic variation is their placement of *wh*- words (*what, who*, and so on) in the sentence. Some languages, such as English, position *wh*- words at the beginning of the sentence, whereas others, such as Chinese, do not. Languages on the whole seem to pattern either like English or like Chinese in this respect. The two options are therefore interpreted as the two possible UG 'settings' for a *wh*-placement parameter. A language learner is said to set the value for such a parameter through direct experience with input from the language being learned. Most parameters are thought to be binary in nature, i.e. they allow only two alternatives, like this one. Parameter setting has been compared with a process of setting electrical switches or filling in the blanks of a form. Thus, parameter setting is the real work of language acquisition. UG provides a structured framework and

all the possible choices, and ensures that the choices will be set correctly.

References

Atkinson, Martin. 1996. Children's syntax. Oxford: Blackwell.
Chomsky, Noam. 1981. Lectures on Government and binding. Dordrecht: Foris.
Cook, Vivian, and Mark Newson. 1997. Chomsky's universal grammar: an introduction. Oxford: Blackwell.
Herschensohn, Julia. 1998. Universal grammar and the critical age. Behavioral and Brain Sciences 21.
Meisel, Jürgen. 1995. Parameters in acquisition. The handbook of child language, ed. by Paul Fletcher and Brian McWhinney. Oxford: Blackwell.
Pinker, Steven. 1994. The language instinct. New York: Harper.

SHEILA DOOLEY COLLBERG

Utterance-Centered Linguistics

When approaching the study of language, a fundamental issue is to determine the basic unit for analysis. This decision will depend on the features of language and its use, which are of most interest to a particular researcher. In the case of utterance-centered linguistics, utterances rather than, for instance, words or sentences are taken as the unit for analysis.

An utterance is any contextualized sequence of naturally occurring language that may be distinguished from other sequences of language spoken or written by the same person or by other people in the same context. Very brief utterances may take the form of far less than a sentence: perhaps just a sigh or a grunt that conveys meaning without words or grammar. Other utterances may be realized by a single word or a grammatically incomplete sentence: for example, *Okay.* or *Got it!* Or utterances may take the form of a whole sentence— *Could you close the window?*—or even a number of sentences in a row (as in an extended request for assistance, with reasons and expressions of anticipatory gratitude). One or more naturally occurring oral or written utterances constitute a discourse. Utterance-centered linguistics studies discourses: component utterances may be analyzed either according to the linguistic features that mark or create coherence among them, or in terms of their embodiment of speech acts. A speech act is an intended meaning conveyed by a speaker or writer. Arguably, the most influential approach to utterance-centered linguistics has focused on speech acts, particularly in oral language.

Although utterances have no fixed grammatical structure, they do have structural requirements with regard to speech acts. Each will convey at least one speech act, sometimes a number of them simultaneously. Consider the following short discourse:

A: *It's twenty minutes past six.*
B: *Sorry. The boss called an unexpected meeting.*

This discourse may be viewed as consisting of three utterances: (1) the comment uttered by *A* and (2) and (3) the response to it uttered by *B*, with each utterance realizing at least one speech act: viz. (1) reproaching, (2) apologizing, and (3) justifying. In this discourse, also, it is noteworthy that whereas *Sorry.* is an incomplete sentence which therefore might be said to have a defective grammatical structure, this is not the only utterance where the hearer's understanding cannot be determined by the grammatical structure of the message alone. For instance, utterance (1)—although grammatically complete—is not accurately interpretable as a reproach only on the basis of its grammatical structure: grammatically, it is just an assertion. Full understanding depends on taking nonlinguistic contextual factors into account as well: what prior arrangement did the two speakers make? and what grounds might *A* have for feeling disappointed or angry if *B* failed to carry through? Clearly, *A* has framed utterance (1) on the assumption that *B* will readily comprehend the reproach through reliance not only on linguistic features but also on contextual knowledge and inferences. And the same assumption underlies *B's* response.

So utterance-centered linguistics views language as a socially contextualized rather than a strictly form-based phenomenon; this focus makes utterance-centered linguistics particularly effective for the study of natural language, since actual language users plainly conceptualize speaking or writing as a process of using utterances to perform speech acts with real-life consequences, not as a process of producing grammatically correct sentences. A key early figure in this tradition was J.L. Austin, the title of whose influential book, *How to do things with words* (1962), reflects his thesis that language use is a form of social action. Utterances are actions in the sense of being attempts to get others to do certain things, to entertain certain thoughts, or to experience certain emotions. Thus, Austin's work drew

1149

attention to the fact that real-life language users would be perplexed or even annoyed if their interlocutors interpreted their utterances only in terms of grammar, which in itself commonly gives no explicit indication of the intended meaning (speech act). For instance, anyone who asked another person in the street, *Do you know where the nearest bus stop is?* would doubtless feel insulted by the reply, *Yes, I know*. To treat an utterance simply as a more or less well-formed grammatical sentence is willfully to miss the point. In the above example, the lost person's utterance is evidently intended as a request for information, not as a mere yes–no question, and so any appropriate response would necessarily include an effort to provide directions.

Analyzing utterances and the utterance structure of discourses requires linguists to develop principles for disentangling separate but sometimes closely similar speech acts, for example *promising* and *threatening*. John Searle (1969, 1979) made important contributions in that respect. Also, he proposed (1975) a useful distinction between direct and indirect speech acts: this contrast has proven to be valuable because it allows analysts to address the way in which, for instance, the above utterance, *Do you know where the nearest bus stop is?* indirectly realizes the speech act of requesting information by directly performing a quite different speech act, a simple yes–no question. A further significant contribution was made by H. Paul Grice (1975, 1981), who suggested a possible set of maxims (guidelines) which—if shared by all speakers in a speech community—could explain the remarkable degree of confidence and accuracy with which speakers formulate and hearers interpret utterances embodying indirect speech acts.

The social orientation of utterance-centered linguistics has had very fruitful implications. Perhaps most profoundly, its willingness to recognize the multiple and indirect functioning of many utterances allows for thought-provoking links with the concept of deconstruction in semiotics, with all its cultural, political, and aesthetic reverberations. Moreover, utterance-centered linguistics has opened the way to important innovations in language teaching methodology and course design, made possible by structuring activities and curricula around utterances rather than sentences.

References

Austin, J.L. 1962. How to do things with words. Cambridge, US: Harvard University Press, and Oxford, UK: Oxford University Press.

Cole, P., and J.L. Morgan (eds.) 1975. Syntax and semantics 3: speech acts. New York: Academic Press.

Grice, H. Paul. 1975. Logic and conversation. In Cole and Morgan.

———. 1981. Presupposition and conversational implicature. Radical pragmatics, ed. by P. Cole. New York: Academic Press.

Schiffrin, Deborah. 1987. Discourse markers. New York and Cambridge, UK: Cambridge University Press.

———. 1994. Approaches to discourse. Cambridge, US and Oxford, UK: Blackwell.

Searle, John R. 1969. Speech acts. Cambridge, UK: Cambridge University Press.

———. 1975. Indirect speech acts. In Cole and Morgan.

———. 1979. Expression and meaning. New York and Cambridge, UK: Cambridge University Press.

Stubbs, Michael. 1983. Discourse analysis: the sociolinguistic analysis of natural language. Chicago: University of Chicago Press, and Oxford, UK: Blackwell.

Widdowson, Henry G. 1978. Teaching language as communication. New York and Oxford, UK: Oxford University Press.

JOHN N. SIVELL

See also **Austin, John Langshaw; Grice, H. Paul; Searle, John; Speech Acts**

V

Variation

What is often known as 'variation theory' in sociolinguistics refers to the research paradigm developed by William Labov in the 1960s for the quantitative study of language variation and change. All variation in speech that cannot be explained on a purely linguistic level was once regarded as random or haphazard. In English, for example, the alternation of *-in'* and *-ing* in words such as *swimmin'* or *swimming* is sometimes called 'free variation', that is, speakers might choose to say *swimmin'* or *swimming* randomly. Variation theory aims to show that this type of variability is by no means random; rather, a statistical and comparative study involving multiple speakers and social contexts can demonstrate that it is correlated with various social and linguistic dimensions. In other words, the goal of variation theory is to explore the systematic nature of variability in language.

To conduct a variation study, investigators first need to select a variable—such as a sound (phonological) segment that tends to vary in pronunciation—and quantify occurrences of variants of this variable in the speech of different individuals in a given community. Take the example of word-final consonant cluster reduction in various English dialects. Words such as *wind, test*, and *pact* are sometimes reduced to *win'*, *tes'*, and *pac'*, respectively. In this case, the two variants of this variable (i.e. word-final consonant cluster such as *-nd, -st*, and *-kt*) are (1) reduced consonant clusters (e.g. *tes'*) and (2) nonreduced consonant clusters (e.g. *test*). Investigators then propose possible factors that might influence word-final consonant cluster reduction. These factors are of two major types. First, there are factors internal to the language system itself.

For instance, the phonological environment that follows the cluster might be an important internal factor—it is possible that word-final consonant clusters are more likely to be reduced when followed by a consonant (e.g. *east coast*) than by a vowel (e.g. *east end*). In addition, there are social or external factors that may favor or disfavor word-final consonant cluster reduction. Age, gender, social class, and contextual style are some of the external factors that are often taken into consideration in variation studies. Investigators might hypothesize that a speaker from a working-class area would be more likely to say *eas'* instead of *east* than speakers from other communities. Similarly, a speaker might say *pac'* for *pact* in casual speech but not in formal speech. Finally, multiple-regression analyses—statistical analyses that determine the significance of multiple factors—are performed. The rule of word-final consonant cluster reduction is not categorical (that is, not all speakers from working-class areas reduce consonant clusters in word-final positions, and no speaker would do so at all times). The goal of the statistical analyses is to determine which factors increase or decrease the likelihood of consonant cluster reduction.

With the use of quantification and statistics, investigators can make accurate statements about fine-grained differences between groups of speakers in a community. When examined at the level of the community, what seems to be random phonetic variation is actually systematically patterned and correlated with various social and language-internal factors. In his study of New Yorkers' speech, for example, Labov (1972) showed that both higher social class and higher

degrees of formality favored conservative (or 'standard') pronunciations. Furthermore, an innovation that occurred in working-class speech was likely to appear in the informal speech of all speakers. In addition to social class and contextual style, Labov showed that variation in New Yorkers' speech was correlated with age as well. In the case of syllable-final *r*, its occurrence was not significant for speakers from all social classes in the two oldest age groups (50–75 and 40–49). In other words, speakers in these age groups did not usually pronounce the *r* in words like *car* and *teacher*. However, the pronunciation was a prestige marker for speakers under 40 years old: speakers from the upper middle class consistently pronounced *r* in syllable-final position more often than those from other social classes, and a regular increase in the use of *r* marked the move from casual to formal speech styles. Examining age as one of the social factors allows investigators to observe language change in 'apparent time' by listening to the speech of different generations in a given community. The use of different variants of a variable shows that language change does not take place abruptly, but rather in small gradations.

Nevertheless, generalizations obtained through statistical techniques are merely descriptive statements. Correlations between variation in language, on the one hand, and linguistic and social dimensions, on the other, need to be explained with solid reasoning. While the relationship between variation in speech and language-internal factors can be accounted for by certain linguistic mechanisms (e.g. word-final consonant cluster reduction can be explained in terms of the simplification process), sociolinguists look to sociology and anthropology for insights into the relationship between language variation and variation in society. Although the notion of social class is still considered a key dimension in sociolinguistic change, John Rickford (1986) and a host of others have argued that social class needs to be modified for non-Western societies. Furthermore, the concept of social network—that is, the sum of relationships that an individual shares with other people—has received increased attention in recent years (see Milroy 1987). Others, most notably Penelope Eckert (2000), advocate an ethnographic approach to the study of language variation and change. Investigators who adopt this approach do not decide a priori which social factors should be examined in relation to language variation; rather, they participate in the everyday lives of the speakers and understand the social dimensions that matter most to members of the community. Instead of relying on ad hoc explanations, they offer more grounded social explanations for linguistic behavior, in that they possess a better understanding of the social meanings associated with linguistic variation in the communities they study.

References

Eckert, Penelope. 2000. Linguistic variation as social practice. Oxford and Cambridge, MA: Blackwell.

Labov, William. 1972. Sociolinguistic patterns. Philadelphia: University of Pennsylvania Press.

Milroy, Lesley. 1987. Language and social networks. Oxford and Cambridge, MA: Blackwell.

Rickford, John. 1986. The need for new approaches to social class analysis in sociolinguistics. Language and Communication 6. 215–21.

Wolfram, Walt, and Natalie Schilling-Estes. 1998. American English. Oxford and Cambridge, MA: Blackwell.

ANDREW WONG

See also **Labov, William**

Vietnamese

Vietnamese is spoken by approximately 65 million of the 80 million inhabitants of the Southeast Asian country of Vietnam. Nearly two million additional speakers of Vietnamese live outside the country, with over a million residing in North and South America, mainly in the United States and Canada, and approximately 400,000 in Europe, particularly in France, Great Britain, Germany, and Switzerland. Speakers of Vietnamese can also be found in Australia, New Zealand, Hong Kong, and Japan. Vietnamese has often played a secondary role to other languages in Vietnam, particularly Chinese, which was the language of Vietnam's ruling class from 111 BC to AD 939, and French, which served as the country's official language from the late nineteenth century until the mid-twentieth century. However, Vietnamese is now recognized as the official language of Vietnam, becoming the medium of instruction in schools in 1945.

Because of the similarities and differences that Vietnamese shares with other Southeast Asian

languages, linguists have had difficulty establishing the genetic affiliation of the language with absolute certainty. Due in part to the great number of Chinese words in the Vietnamese lexicon, Vietnamese was at one time considered to be a member of the Sino-Tibetan family. More recently, linguists have attributed this similarity in word stock to the long domination of Vietnam by the Chinese and have explored similarities between Vietnamese and the members of other language families. French scholar Henri Maspéro (1912) proposed that the genesis of Vietnamese resulted from the fusion of a member of the Mon-Khmer family, such as Khmer (formerly Cambodian), and a Tai language, such as Thai. Ultimately, Maspéro identified Vietnamese as a Tai language because of its use of tone, a characteristic feature of Tai languages that is uncharacteristic of Mon-Khmer languages. A number of other linguists, however, have argued that Vietnamese originated as a Mon-Khmer language that adopted a tonal system through contact with speakers of Thai. French botanist–linguist André Haudricourt (1954) argued that the phenomenon of tonal genesis in Vietnamese was an evolutionary process in which the language acquired three tones by the sixth century through language contact, and doubled this number by the twelfth century to the six tones currently used in the Hanoi dialect of Vietnamese. This account, coupled with the observation of the similarity between Vietnamese and Muong by Pryzluski (see Meillet and Cohen 1924), has convinced many linguists that both Vietnamese and Muong belong to the Mon-Khmer phylum in the Austro-Asiatic family.

Although linguists have repudiated anything more than a distant genetic relationship between Vietnamese and Chinese, the two languages have a number of similarities. The most apparent among these is the great number of Chinese words in the Vietnamese lexicon, many of which were borrowed during the Chinese domination of Vietnam from 111 BC to AD 939. During this period, Chinese was not only used among many of the officials of Vietnam, but also as a medium of education; as a result, the language of law, politics, history, medicine, science, and technology in Vietnamese include a great number of borrowings from Chinese. While some of these words have been modified to such an extent by Vietnamese speakers that they can hardly be recognized as Chinese anymore, others are referred to as 'Sino-Vietnamese' because they are pronounced as they would have been in ancient Chinese. Vietnam's proximity to populations of Malay and Thai speakers has resulted in the Vietnamese lexicon including borrowings from these languages; Vietnamese has also borrowed words from French, English, and Russian as a result of colonization, war, and political alliance.

Vietnamese and Chinese share a number of other characteristics leading to the false impression that the two languages are closely related. One of the most significant of these characteristics is that both languages are considered tonal because they recognize tone as phonemic, i.e. tone makes a difference in meaning. For example, in the Hanoi dialect of Vietnamese, which recognizes six different tones, the word phonetically realized as [ma] has six different meanings depending on which of the six tones (high, low-falling, high-rising, creaking-rising, dipping-rising, and constricted) is used. In addition to the explanation of tonal genesis by Haudricourt, linguists have attributed this commonality to contact between speakers of Vietnamese and Chinese, particularly the Chinese dialects spoken in the southern provinces near the border of Vietnam, which tend to have more tones than other Chinese dialects.

Like other languages spoken in Southeast Asia, Vietnamese is a morphologically isolating language, a typology that Vietnamese also shares with Chinese. Characterized by the absence of bound inflections, isolating languages contrast with synthetic languages, which use affixes on words to indicate such things as number, tense, and agreement. For example, by affixing the bound morpheme /s/ to a noun or /ed/ to a verb, speakers of English respectively show plurality and past tense. In an isolating language like Vietnamese, however, words and morphemes tend to have a one-to-one relationship, so that when Vietnamese speakers want to show plurality, they must use a separate word or grammatical particle elsewhere in the sentence. As in Chinese, there tends to be a one-to-one relationship between syllables and morphemes in Vietnamese, which has led some scholars to deem Vietnamese as monosyllabic. Other scholars, however, have taken issue with this characterization and have pointed out that the Vietnamese lexicon includes many words that are disyllabic, as well as words that have three or four syllables (Hannas 1997:76–77). The perception that Vietnamese is monosyllabic has also been attributed to the Vietnamese orthographic practice of putting a space between every syllable of a word.

Although it lacks inflectional morphology, Vietnamese employs derivational morphology as a means for creating new words. Vietnamese uses prefixing and suffixing, which are two processes of affixation used productively in many of the world's languages, including English; however, in contrast to many other Southeast Asian languages, Vietnamese does not use infixing, a type of affixation in which a morpheme is inserted into the root of a word. Another very productive process of deriving new words in Vietnamese is compounding, not only as a process of creating new nouns but in the derivation of new verbs and adjectives. Reduplication, which is the process of deriving a word

by copying all or part of a word and affixing it to itself, is yet another morphological process that is used often in Vietnamese and other Southeast Asian languages, as well as Chinese. Like many of the world's languages that use reduplication productively, Vietnamese commonly uses reduplication to pluralize and intensify the original semantics of a word, as well as to show reciprocity. Reduplication is also commonly used in onomatopoetic expressions in Vietnamese, as is also the case in other languages of the world.

The basic syntactic structure of Vietnamese is subject–verb–object (SVO). Because it is an isolating language, Vietnamese relies heavily on word order to show the grammatical roles being played by various arguments in a sentence. Vietnamese also uses special markers to indicate such things as the tense of verbs and the number and gender of nouns. In Vietnamese, a tense marker is placed before the verb that it modifies, and negative markers also appear in preverbal position. Adjectives come after the nouns they modify in Vietnamese. Content questions, or questions that demand an answer besides yes or no, are formed by inserting an interrogative word in the position in the sentence in which the unknown material should be, leaving the sentence in its basic SVO order. Although SVO is the basic word order in Vietnamese, object–subject–verb (OSV) order is also commonly used when a speaker moves an object to the beginning of a sentence to give it emphasis over the subject, a process known as topicalization.

The system of personal pronouns in Vietnamese is rather complex due to the use of pronouns not only as substitutes for the name of a person or group of persons, as they are in English, but also as a way of marking the status of participants and referents in conversations. Based largely on factors such as age and familiarity, the use of various pronouns to mark status requires speakers to evaluate whether they are superior or inferior to those with whom they are speaking, and this decision affects which first-person pronouns they use, as well as second-person pronouns. As is the case in other cultures that use languages with similar pronoun systems, there are social consequences for choosing the wrong pronouns in Vietnamese. The pronoun system also has two forms for the first-person plural pronoun: one to be used when speakers wish to include their addressees in the group for which they are using the pronoun (inclusive) and another if the addressees are not included (exclusive).

Like other Asian languages, Vietnamese also has a complex system of kinship terms, or terms that speakers use to address and to refer to members of their own family, which reflects the extended and patrilineal nature of the family structure. In speaking to members of their immediate family, children must use various

first-person pronouns depending on whether they are addressing their parents, elder siblings, or younger siblings. In using the second person, children must use pronouns reflecting the distinction in gender when addressing their parents, a distinction that is preserved when children address elder brothers and sisters. This distinction is not maintained when parents address children or when elder siblings address younger siblings. As in the use of personal pronouns, the choice of an inappropriate kinship term may have consequences. Apparently borrowed from the South Chinese, one custom that is commonly used among the lower classes of Vietnam is the use of *kin numeratives*. A device used for designating grown-up children, a common pattern is to call the eldest child 'Number One', the second child 'Number Two', and so on. An interesting variation of this occurs in southern Vietnam, where 'Number Two' is used to refer to the eldest son or daughter, with speculation that the Vietnamese word for 'Number One' was reserved for the mother because she is considered the father's first child (Dinh-Hoa 1980:32).

Three different orthographic systems have been used during the history of writing in Vietnamese. The first was *chu nho*, which was borrowed directly from the Chinese and was the official written language from as early as the eleventh century until the nineteenth century. The next was *chu nom*, a writing system based on the Chinese system but altered to phonetically represent spoken Vietnamese. The invention of this system is generally attributed to Han Thuyen, a famous Vietnamese poet of the thirteenth century. The modern spelling system is *quoc ngu*, a Roman alphabet devised by Catholic missionaries from Europe to translate scripture into the language in a way that would make sense to the Vietnamese, as well as to the missionaries who read aloud from it. During French domination of the country from the late nineteenth century to the mid-twentieth century, the French used *quoc ngu* to introduce the Vietnamese to French language and literature. Although many have praised the success of the Vietnamese in replacing the Chinese writing system with a more phonetic alphabet, others have pointed out that the extensive use of diacritics to indicate tones in *quoc ngu* creates a system nearly as complicated as the one it was intended to replace (Hannas 1997:75).

Vietnam is characterized by three dialect areas: northern, central, and southern, perhaps reflecting the division of Vietnam into three states by the French during their domination. The northern dialect area includes Hanoi, the capital city of Vietnam. Ho Chi Minh City (formerly Saigon) is the largest city in Vietnam and is in the southern dialect region. The differences between the three dialects are mainly at the levels of pronunciation and vocabulary and are not great enough to cause

1154

much difficulty in communication among speakers of different dialects. Although Vietnam does not have a spoken standard, with each locale maintaining its own dialectal flavor, the dialect spoken in Hanoi is the dialect used most often in the media.

References

Dinh-Hoa, Nguyen. 1980. Language in Vietnamese society: some articles by Nguyen Dinh-Hoa. Carbondale, IL: Asia Books.

———. 1997. Vietnamese. Amsterdam: John Benjamins.

Gregerson, Kenneth J. 1969. A study of middle Vietnamese phonology. Saigon: Summer Institute of Linguistics.

Hannas, Wm. C. 1997. Asia's orthographic dilemma. Honolulu: University of Hawaii Press.

Haudricourt, André-Georges. 1954. De l'origine des tons in viêtnamien. Journal Asiatique 242. 68–82.

Luong, Hy V. 1990. Discursive practices and linguistic meanings: the Vietnamese system of personal reference. Amsterdam: John Benjamins.

Maspéro, Henri. 1912. Études sur la phonétique historique de la langue annamite. BEFEO 12. 1–127.

Meillet, A., and Marcel Cohen. 1924. Les Langues du Monde. Paris: Centre national de la recherche scientifique.

Ngo, Binh Nhu. 1999. Elementary Vietnamese. Boston: Tuttle.

Thompson, Laurence C. 1965. A Vietnamese grammar. Seattle: University of Washington.

LAMONT ANTIEAU

See also **Austroasiatic; Chinese (Mandarin); Khmer and Mon-Khmer Languages; Southeast Asia; Thai and Tai Languages**

Visual Word Recognition

Visual word recognition refers to the process by which readers identify written words. Recognizing a word enables a reader to use the word's phonology to read aloud or to use the word's semantics to determine its meaning. Skilled reading is marked by the ability to recognize words quickly and with relatively little effort.

There are two major approaches to modeling cognition in general, and word recognition in particular. Cognitive neuropsychologists of the nineteenth century assumed that the language processing system was modular and that component structures and processes could be represented as a box-and-arrow diagram. This approach fell out of favor in the first half of the twentieth century, but remains popular in the twenty-first century. The revival is bracketed by two seminal articles in *Psychological review:* Morton's (1969) logogen model and Coltheart's (2001) dual route cascaded (DRC) model.

That portion of the DRC concerned with reading aloud can be used to provide an example of the modular approach. The model assumes two principal pathways by which a printed word can be read aloud. Both routes are initiated by a common process of orthographic analysis, which leads to the identification of the word's constituent letters. The lexical route uses the letter identities to activate an entry in a mental dictionary called the orthographic input lexicon. In turn, the pathway from the orthographic input lexicon to the phonological output lexicon activates the word's phonology. The other major route processes the letters in left-to-right order and applies a set of grapheme–phoneme rules. For example, the grapheme consisting of the first two letters of SHEEP is translated as the phoneme that sounds like 'sh'. Both of these routes compute a pronunciation that is represented as a sequence of phonemes. The eventual response is based on combining the outputs of both routes in a response buffer.

The lexical route is needed in order to correctly pronounce words that have an irregular spelling–sound correspondence such as PINT because the phonology generated by the rule route favors a pronunciation that would rhyme with the regular words HINT and MINT. The rule route is needed in order to model the ability of humans to easily read aloud nonwords like MAVE. In this case, the lexical route would have no entry for MAVE and could not generate a pronunciation other than by analogy to similar real words. This division of labor allows dual-route models to offer compelling explanations for different types of acquired dyslexias that together form a double dissociation. For example, damage to the lexical route should lead to specific difficulties with irregular words like PINT (surface dyslexia), but preserve the ability to read nonwords and regular words. On the other hand, damage restricted to the rule route should lead to specific difficulties in reading nonwords (phonological dyslexia), but preserve the ability to read both regular and irregular words.

The DRC is also a computational model, meaning that it has been implemented as a computer program that performs reading tasks by using the same structures and processes hypothesized for human readers. The reading tasks studied most often are reading aloud
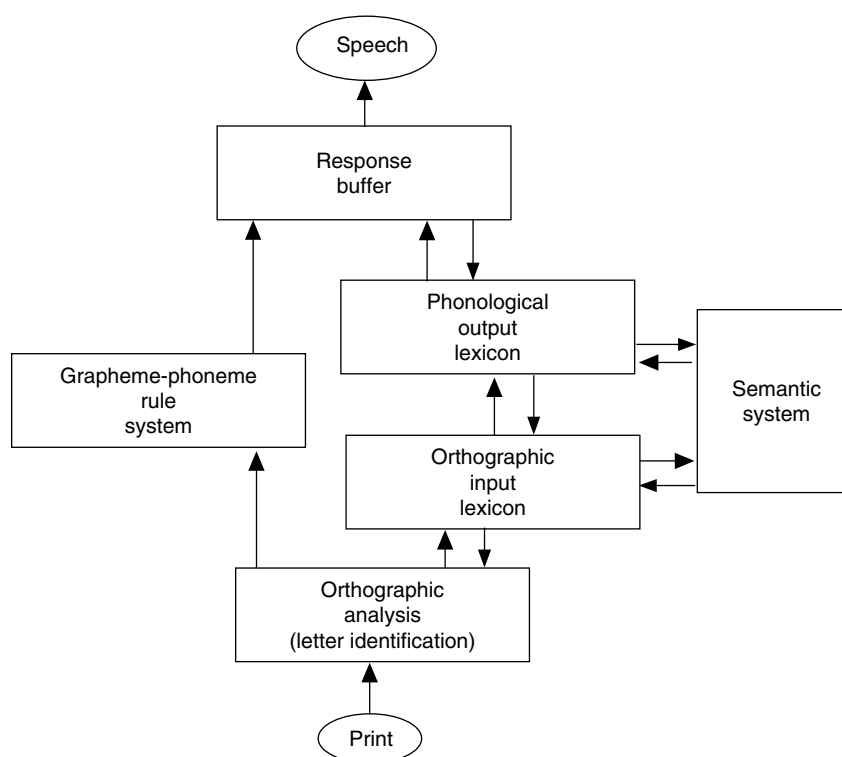
Figure 1
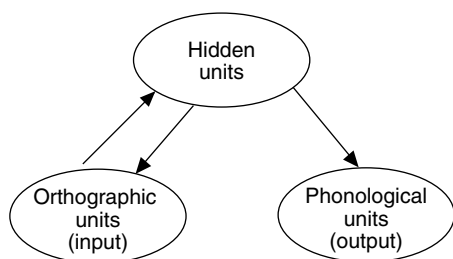
Figure 2

and the lexical-decision task. In the latter, participants are asked to make speedy responses as to whether a letter string is a word or a nonword. DRC can simulate the speed and accuracy of human readers in response to many different characteristics of words. These include spelling–sound regularity, frequency of occurrence, and the orthographic similarity of a target word to other words.

One heavily researched phenomenon in word recognition that the DRC has not simulated is semantic priming. Semantic priming occurs when the response to a target word like DOCTOR is faster when preceded by a related prime like NURSE. In theory, DRC could account for semantic priming because, in addition to the direct pathway from the orthographic input lexicon to the phonological output lexicon, there

is also a mediated pathway through a semantic system that contains semantic representations of each word. However, the semantic pathway has not been implemented in the computational model because developing realistic semantic representations is more difficult than representing orthography and phonology. However, powerful new tools like latent semantic analysis and other high-dimensional representations based on the co-occurrence of words in very large corpora of text make it likely that computational models will soon be able to address more adequately the very important role of semantics in word recognition.

The second major approach used to model word recognition uses learning algorithms in conjunction with networks that have distributed representations. The seminal model of this type was Seidenberg and McClelland's distributed developmental model (DDM, 1989). Using the DDM as an example and reading aloud as the task of interest, this computational model is based on a network with three layers. The input units represent orthography, but that representation is distributed across a large set of nodes. This means that individual nodes do not symbolize individual letters. Rather, a given node is sensitive to many letter sequences and, conversely, a letter in a specific input location will partially activate many different input nodes. The third, or output, layer codes phonology with a similar type of distributed representation.

The middle hidden layer has no symbolic representations at all. Each node in one layer is connected to every node at the next layer. Each connection has a weight that is initially set at a random value, but during training the back propagation learning algorithm progressively adjusts the weights such that each input (printed word) more and more closely approximates the correct output (pronunciation). In the DDM, both irregular words and nonwords can be read aloud by using a single pathway from input (orthography) through the hidden layer to output (phonology). In contrast to dual-route models, it does so without the explicit representation of grapheme–phoneme rules or the explicit representation of words in a mental lexicon. The DDM and its successors have also enjoyed considerable success in simulating the performance obtained with people in reading tasks.

References

Besner, Derek, and Glyn Humphreys (eds.) 1991. Basic processes in reading: visual word recognition. Hillsdale, NJ: Lawrence Erlbaum Associates.

Coltheart, Max, Kathleen Rastle, Conrad Perry, Johannes Ziegler, and Robyn Langdon. 2001. DRC: A dual route cascaded model of visual word recognition and reading aloud. Psychological Review 108(1). 204–56.

Frost, Ram, and Leonard Katz (eds.) 1992. Orthography, phonology, morphology, and meaning. Amsterdam: North-Holland.

McClelland, James, and David Rumelhart. 1988. Explorations in parallel distributed processing. Cambridge, MA: MIT Press.

Morton, John. 1969. Interaction of information in word recognition. Psychological Review 76. 165–78.

Seidenberg, Mark, and James McClelland. A distributed, developmental model of word recognition and naming. Psychological Review 96. 523–68.

KENNETH R. PAAP

Voice

'Voice' is a technical term for the description of sentence structure. It encompasses a range of grammatical constructions that differ in the correlation between the semantic and grammatical function of particular sentence elements. The English sentence *John kissed Mary*, for example, has two noun phrases: *John* and *Mary*. The first, the subject, performs the action described by the sentence and is thus interpreted as the 'agent'. The second, the object, undergoes the action and is thus interpreted as the 'patient'. In an English active sentence such as this one, the grammatical function 'subject' thus correlates with the semantic role 'agent', whereas the 'object' fulfills the semantic role of 'patient'. In languages such as English, 'active' is the basic voice, because it appears with all verbs and does not require any specific marking on the words involved (e.g. *was kissed*). These languages typically also have 'passive' and 'middle' constructions, which are discussed below.

There are many languages (so-called ergative languages), however, in which sentences where the patient surfaces as subject are the basic ones. A cross-linguistically valid definition of voice is thus difficult, and any characterization of voice in terms of information sequencing or in terms of cognitive processes raises a number of problems, because the voice systems of different languages differ enormously from one another.

A passive construction or sentence is one in which a somehow marked form of the verb has a nonagent (typically a patient) subject, whereas the agent may or may not appear as an object introduced by a preposition (e.g. *Mary was kissed [by John]*). The passive has traditionally been analyzed as the result of a grammatical operation (passivization); the active sentence used to be regarded as the basic one, which would somehow need to be changed to become a passive sentence. Passive sentences were considered to be synonymous paraphrases of their active counterparts. *John kissed Mary* and *Mary was kissed by John* were regarded as essentially equivalent utterances. Since the 1980s, however, several studies have taken into consideration that the passive may have different connotations in context; i.e. the choice between an active and a passive sentence is more than just a random stylistic one. Rather, it is seen as a matter of relative prominence of agent and patient—passive and active differ in terms of information packaging or sequencing, or in terms of foregrounding of information. Because the subject is usually the topic of the sentence, a passive sentence is thus argued to foreground the patient role. Although supported by psycholinguistic experiments, this view of passive constructions has been widely challenged, and it has been claimed that the main function of passive sentences is not the foregrounding of the patient but rather the defocusing or backgrounding of the agent. This would account for the fact that in a significant number of cases the passive voice does not signal that

the patient is topical, but simply indicates that the agent is not important. A case in point is the so-called impersonal passive (e.g. German *Es wurde hier getanzt,* 'There was dancing here'). Because in impersonal passives there is no patient role, patient topicalization cannot be considered as the function of these passives.

Some verbs, such as English *sink,* appear in pairs such as *The enemy sank the ship* vs. *The ship sank.* The object of the first clause corresponds to the subject of the second. The verb in the second clause would thus be expected to occur with passive marking, which, however, is not the case. These verbs have sometimes been labeled 'middle verbs' because they seem to break out of the active/passive dichotomy. Other constructions typically referred to as 'middle' are reflexive sentences. An example would be the Spanish sentence *Juan se ve,* 'John sees himself'. The primary function of this construction is to encode the identity of agent and patient. The argument that this is somewhat in between active and passive has been extended to other functions or meanings of the reflexive marker, such as the encoding of an action performed on the subject's body, as in the Spanish *Juan se peina el pelo,* 'John combs his hair'; the expression of situations that occur entirely within the subject's sphere, as in sentences with the Spanish verb *sentirse,* 'feel' (literally 'feel-self'); or even the expression of passive meaning, as in Spanish *aquí se habla Español,* 'Spanish is spoken here'. In some languages, the opposition active/middle is actually encoded in alternating verb forms. This can be observed in Sanskrit and Greek (both ancient and modern), both members of the Indo-European family, and in languages such as Tula (a Niger-Congo language) and Tamil (a Dravidian language). In these languages, the middle voice usually denotes actions that are performed on the body or on a part of it, as well as spontaneous actions, i.e. events presupposing no causal agent.

Thus, many languages distinguish between active and passive voice and may also include a middle voice. Ergative languages approach the same issue from the opposite side and seem to focus on the patient rather than the agent. To further complicate the issue, in some languages, the question of whether or not the subject of the sentence is an agent or not becomes altogether secondary. These languages, for instance, Philippine languages or native American Algonquian languages, concentrate instead on 'empathy'. Empathy is an all-encompassing complex system involving features such as animacy, (perceptive) salience, (potential) agency, whether or not something has been mentioned before, etc. 'Inverse voice' contrasts here with 'direct voice'. The inverse voice indicates that the agent ranks lower on the empathy scale than the other role, whereas the direct voice is used when the agent is higher on the empathy scale. In both inverse and direct clauses, the agent usually has subject properties; for example, it agrees with the verb. Talmy Givón (1994) distinguishes between semantic and pragmatic inverse systems. In the latter, the choice between a direct or an inverse clause is a matter of which of the two participants is more salient in the specific discourse. However, in semantic inverse systems, the decision between inverse and direct voice depends on more objective relations, such as animacy: if the subject refers to an animate entity (e.g. an animal), while the object refers to an inanimate one (e.g. a rock), the direct voice would be used, because the animate subject is higher on the empathy scale than the inanimate object.

References

Andersen, Paul Kent. 1991. A new look at the passive. Frankfurt: Lang Verlag.

Blevins, James P. 2003. Passives and impersonals. Journal of Linguistics 39.

Cornelis, Louise H. 1997. Passive and perspective. Amsterdam and Atlanta, GA: Rodopi.

Dixon, R.M.W. 1994. Ergativity. Cambridge, UK, New York: Cambridge University Press.

Fox, Barbara, and Paul J. Hopper (eds.) 1994. Voice: form and function. Amsterdam and Philadelphia, PA: John Benjamins.

Givón, Talmy. 1981. Typology and functional domains. Studies in Language 5.

——— (ed.) 1994. Voice and inversion. Amsterdam and Philadelphia, PA: John Benjamins.

Haspelmath, Martin. 1990. The grammaticization of passive morphology. Studies in Language 14.

Jaeggli, Osvaldo. 1986. Passive. Linguistic Inquiry 17.

Keenan, Edward L. 1985. Passive in the world's languages. Language Typology and Syntactic Description, Vol. 1, ed. by Timothy Shopen. Cambridge: Cambridge University Press.

Kemmer, Suzanne E. 1993. The middle voice: a typological and diachronic study. Amsterdam and Philadelphia, PA: John Benjamins.

Kirsner, Robert S. 1976. On the subjectless 'pseudo-passive' in standard Dutch and the semantics of background agents. Subject and topic, ed. by Charles N. Li. New York: Academic Press.

Klaiman, M.H. 1992. Middle verbs, reflexive middle constructions, and middle voice. Studies in Language 16.

Myhill, John. 1997. Toward a functional typology of agent defocusing. Linguistics 35.

Perlmutter, David M., and Paul M. Postal. 1977. Toward a universal characterization of passivization. BLS 3. Berkeley, California.

Sansò, Andrea. 2003. Degrees of event elaboration. Passive constructions in Italian and Spanish. Milan: Francoangeli.

Shibatani, Masayoshi. 1985. Passive and related constructions: a prototype analysis. Language 61.

Siewierska, Anna. 1984. The passive. A comparative linguistic analysis. London: Croom Helm.

ANDREA SANSÒ

See also **Functional Approaches; Generative Grammar; Syntax; Thematic Structure**

Vowel Harmony

Vowel harmony is a well-known but not yet well-understood phenomenon found in many languages native to Eurasia, Africa, and, to a lesser extent, other places. Hungarian, Finnish, and Turkish number among the best-known cases of harmony. The Uralic and Altaic language families, spoken across a vast geographic area of Eurasia from Korea to Turkey, exhibit vowel harmony to varying degrees in most of their member languages (Hungarian, Finnish, Turkish, Mongolian, Manchu, and so forth). Harmony is not found in any familiar Indo-European languages (e.g. Greek, French, or English).

Simply stated, harmony is when vowels within a word are required to resemble one other in terms of some property. For example, in the Tuvan word *tooruksug,* meaning 'smelling like a pine cone', all vowels are 'labial' (pronounced with rounded lips). Tuvan has an ample supply of vowels that do not use lip-rounding, yet the vowels here—taking their cue from the first vowel—are all rounded.

There are several basic types of harmony, each named for the speech organs involved. These include palatal/velar (also called 'backness') harmony, labial (also called 'rounding') harmony, and tongue root (also called ATR) harmony. Palatal harmony requires vowels to be alike in terms of whether they are pronounced with the tongue body toward the front of the oral cavity (close to the palate) or toward the back (close to the velum). Labial harmony requires vowels to be alike in having lip rounding. Tongue root harmony requires vowels to be alike in whether the tongue body is pushed toward the front of the mouth or pulled toward the back.

A language may have more than one harmony system. Turkish has palatal and labial harmony, whereas Classical Manchu had tongue root and labial harmony. Two separate harmony systems, when present, may operate in an independent or semi-independent manner.

Not all vowels participate equally in harmony. Typically, some restrictions are placed on which vowels cause harmony and which ones obey it. Such restrictions usually refer to some feature other than the harmonic one. In Tuvan, for example, any labial vowel can trigger labial harmony (forcing adjacent vowels to also be labial), but only high vowels (vowels that are pronounced with more tension in the tongue and mouth, such as in *seek* or *plume*) undergo labial harmony (vowels that obey harmony are underlined herein):

Tuvan

Low vowels (pronounced with less tension, as in *long*) do not obey labial harmony:
  xol-da 'in (the) hand'
  xöl-de 'in (the) lake'
  High vowels must obey labial harmony:
  xol-u̲ 'his hand'
  xöl-ü̲ 'his lake'

Vowel harmony may be thought of as a restriction on which vowels may occur together within the same word. In Turkic languages, all vowels can be divided into two natural classes: those pronounced with the tongue in the back part of the oral cavity, and those pronounced with the tongue toward the front. In a Turkic language with eight vowels, four of them will belong to the front (palatal) class and four to the back (velar) class. For a typical Turkic vowel inventory, see below:

	front (palatal)	*back (velar)*
High	i ü	ɯ u
Low	e ö	a o

(labial vowels are ü, u, ö, o)

Within any single word (including any affixes added to that word), vowels from the front and back classes do not mix:

Tuvan

is	'footprint'
is-te̲r-i̲vi̲s-te̲	'in our footprints'
aas	'mouth'
aas-ta̲r-u̲vu̲s-ta̲	'in our mouths'

Harmony applies only within a word. No harmony system allows harmony to carry across a longer span, such as two distinct words or an entire sentence. At the level of word structure, harmony systems may be divided into two basic types, depending on what element controls harmony. In 'root-controlled' harmony, vowels in a word root determine the quality of vowels appearing in affixes. In 'dominant–recessive' harmony, neither roots nor affixes take precedence. Instead, if any vowel belonging to the dominant class appears anywhere in the root or a suffix, it forces all other vowels in the word to shift over to that class. A word may contain weak (recessive) vowels only if no strong vowels are present.

No harmony system is exceptionless: typically, a language will have some nonharmonic native words or affixes. A language that has been in long-term contact

with a nonharmony language (e.g. Russian or Farsi) may have a large supply of nonharmonic loan-words.

Nonetheless, these disharmonic elements may participate productively in harmony. Usually, the final vowel of a nonharmonic word determines the harmony of the following suffix vowels:

Classical Manchu (disharmonic words)

age	'prince'
age-s<u>e</u>	'princes'
dahema	'uncle'
dahema-s<u>a</u>	'uncles'

How is harmony related to the mechanics of speaking? Vowel harmony makes some sense when viewed in terms of natural movements of the speech organs. For example, when you round your lips to pronounce 'oo', this gesture may carry over onto the next vowel. This tendency might account for common slips of the tongue in English, for example, saying *kangaroo moot* when you meant to say *kangaroo meat.* Physical speech mechanisms provide a natural foundation for harmony. But they do not explain why certain languages (and not others) adopt harmony as a fundamental organizing principle of their sound systems. Harmony likely traces its origins to natural speech mechanics, but it has evolved into a more abstract way of organizing the sounds of particular languages. Thus, harmony can work at cross-purposes with the mechanics of speech— it may impose conditions that do not obviously facilitate either speech production or perception.

For example, in Tofa, if a word contains [j] (the sound written with a 'y' in English), all vowels next to [j] must be pronounced in the front of the mouth, as if they belonged to the palatal class. However, for purposes of harmony, the speaker may ignore the fronting effect that [j] has on some vowels, considering them to be back vowels despite their palatal quality. Any suffix the speaker adds will thus contain a velar vowel:

Tofa

xöjen-d<u>a</u>	'in (the) cliff'

In this case, harmony clearly does not facilitate pronunciation, because Tofa speakers are not accustomed to mixing palatal and velar vowels within a word. But it does serve to maintain the integrity (at an abstract level) of harmony as a system.

Besides ease of pronunciation for the speaker, what other advantage might harmony offer in language design? Some linguists have argued that harmony helps to make clear to the hearer the boundary between one word and the next. But this can be true only if two adjacent words belong to different harmony classes. It has also been suggested that harmony increases the likelihood that the hearer will correctly identify certain vowels. If a speaker notices that the vowel of the first

syllable is palatal, he or she can be fairly certain that each following vowel also belongs to the palatal set. This reduces by one half the range of potential vowels he or she must listen for in succeeding syllables, and it may lessen the cognitive processing required.

Many harmony languages have vowels that fail to participate fully or at all. Instead, they may either obstruct the harmony pattern or remain invisible to it. So-called opaque vowels block the current harmony pattern and start a new pattern. The Turkish suffix *-gen* (corresponding to English *-gon*, as in *octagon*) contains a front (palatal) vowel that never alternates to obey harmony. Vowels following *-gen* must be palatal, even when the prevailing pattern in the word is velar:

Turkish

alt ɯ	'eight'
alt ɯ-gen	'octagon'
alt ɯ-gen-l<u>e</u>r	'octagons'

So-called transparent or neutral vowels allow harmony to pass through them, without being affected and without blocking it. In Karelian (a Uralic language), backness harmony dictates that if the first syllable contains a front vowel, all the following syllables contain front vowels. Likewise, if it contains a back vowel, the following syllables have back vowels. Exceptions are the front vowels [e] and [i], which remain indifferent to harmony and may appear anywhere.

Most harmony languages tolerate some disharmony, which can result from internal processes or from borrowing foreign words. Internally, consonants may interfere with the harmony system, imposing their own conditions that override it. Syllable structures may also impose limitations on harmony: in Turkmen, for example, labial harmony does not affect vowels in word-final syllables containing a vowel. In Turkish, the presence of a velar consonant [g] may force an adjacent vowel to be velar, even though the other vowels in the word are palatal. Modern Turkish also has a high percentage of nonharmonic loan-words in its lexicon: *taksi* 'taxi', *kitap* 'book'. Some harmony languages alter borrowed words to render them more harmonic.

The Turkic languages range from almost perfectly harmonic to hardly harmonic at all. These languages illustrate the possibilities of change and evolution in harmony. The plural suffixes in Old Turkic and two of its daughter languages, Turkish (highly harmonic) and Uzbek (barely harmonic), show two possible paths along which harmony can proceed (see below).

	back suffix	front suffix
Old Turkic (harmonic)	-l<u>a</u>r	-l<u>e</u>r
Turkish (harmonic)	-l<u>a</u>r	-l<u>e</u>r
Uzbek (not harmonic)	-lar	-lar

Harmony as a pattern can best be understood by considering those in which it fails to apply as well as those in which it applies. Many unanswered questions remain in understanding vowel harmony. Dozens of languages known to have harmony remain largely undocumented, and new types will certainly be discovered.

References

Anderson, Gregory D.S., and K. David Harrison. 1999. Tyvan [Tuvan]. Munich: Lincom Europa.

Clark, Larry. 1998. A reference grammar of Turkmen. Wiesbaden: Harrasowitz Verlag.

Johannson, Lars, and Eva Csato (eds.) 1999. Turkic languages. London: Routledge.

Kornfilt, Jaklin. 1997. Turkish. London: Routledge.

Li, Bing. 1996. Tungusic vowel harmony: analysis and description. The Hague: Holland Academic Graphics.

Sinor, Denis (ed.) 1988. The Uralic languages. Leiden: E.J. Brill.

Vago, Robert M. (ed.) 1980. Issues in vowel harmony. Amsterdam: John Benjamins.

DAVID K. HARRISON

W
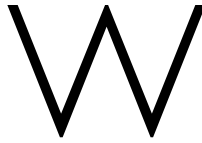
Wackernagel, Jacob

Wilhelm Wackernagel, Jacob's father, originally from Berlin, moved to Basle in 1833, where he was appointed professor for German Studies at the local university. The Wackernagel family developed a very close link with the city of Basle: Jacob's brother Rudolf, who worked for the city archives, wrote a history of Basle reaching back to the sixteenth century, and Jacob lived in Basle and taught at the University of Basle until his death, with only one interruption, during which he taught at the University of Göttingen, in the years preceding World War I. Jacob's life was entirely devoted to studying and teaching classical philology and the ancient Indo-European languages. He had his first appointment at the age of 24; his teacher of Greek was Friedrich Nietzsche.

Wackernagel can be considered a member of the second generation of the 'Neogrammarians', a school of linguists interested in studying how languages change over time. Among his contemporaries, he occupies a special position because of his interest in sentence structure, something that had been investigated to a much lesser extent than pronunciation and word structure. A survey of the courses he gave in his years as professor shows that syntax was a recurring topic already in the 1880s; in 1904 Wackernagel was offering a course in historical syntax, a topic which can still sound innovative today.

In his 1892 essay *Über ein Gesetz der indogermanischen Wortstellung (On a law about Indo-European word order)*, Wackernagel pointed out that in the ancient Indo-European languages a number of particles, conjunctions, pronouns, and verbs took a peculiar position in the sentence, nowadays known as the 'Wackernagel position' or

'P2', because it is situated immediately after the first word with intonational emphasis (accent). All words in P2 share the characteristic that they are unaccented (or weakly accented). This tendency is most visible in Sanskrit and Homeric Greek, but Wackernagel found its traces in other Indo-European languages as well, and he surmised that Proto-Indo-European, the hypothesized ancestor language of all Indo-European languages, followed the same pattern. Thus, Wackernagel reconstructed Proto-Indo-European as having essentially a similar word order to Modern High German, having the tensed verb obligatorily in second position in main clauses and in final position in subordinate clauses.

Wackernagel's hypothesis was received with varying degrees of approval by his contemporaries; especially the idea that the position of the tensed verb in Proto-Indo-European was the same as in German did not meet general consent. However, the fact that various types of unaccented item were placed in P2 found a striking confirmation when Hittite was finally deciphered in 1916, for unstressed pronouns and particles follow the pattern described by Wackernagel virtually without exception in this language.

In the second half of the twentieth century, progress in the study of unaccented particles (usually referred to as 'clitics') and knowledge of previously unknown non-Indo-European languages has demonstrated that the Wackernagel position is not a unique feature of Indo-European. Being unstressed, clitics need what is commonly called a phonological host, i.e. an accented word with which they can form a phonological unit. P2 is one of the possible positions for clitics, found in e.g. some Australian and Uto-Aztecan languages.

Wackernagel's other major accomplishment in the field of syntax is constituted by the two volumes of lectures on syntax (*Vorlesungen über Syntax*), 1926–1928, in which the scholar addressed several problems connected with the syntactic behavior of nouns and verbs in the Indo-European languages (mostly based on Latin, Greek, and Germanic). Wackernagel's approach is typical of historical linguistics in the late nineteenth to early twentieth century, insofar as—from today's perspective—the word is still the dominant unit of analysis. In Wackernagel's times, linguists were in the process of developing distinctive theories of word and sentence structure, but word-related issues were still defining and delimiting syntactic problems. Although he announced a further volume on sentence structure, Wackernagel never wrote it. Since the newly individuated field of syntax could not yet rely on a theoretical framework, Wackernagel's syntactic work remains descriptive, his insights mostly deriving from his deep knowledge of languages and texts.

Biography

Jacob Wackernagel was born in Basle in 1853. He studied Classical Philology in Basle, and was then appointed in the same university, where he taught from 1877 to 1902. He was appointed in Göttingen in 1902–1915. Then he returned to Basle and resumed his former position, which he held until 1937. He died in Basle on May 22, 1938.

References

Collinge, N.E. 1985. The laws of Indo-European. Amsterdam and Philadelphia: Benjamins.

Eichner, Heiner, and Helmut Rix (eds.) 1990. Sprachwissenschaft und Philologie. Jacob Wackernagel und die Indogermanistik heute. Wiesbaden: Reichert.

Hale, Ken. 1983. Warlpiri and the grammar of non-configurational languages. Natural Language and Linguistic Theory 1(1). 5–47.

Halpern, Aaron, and Arnold Zwicky (eds.) 1996. Approaching second: second position clitics and related phenomena. Chicago: CSLI Publications.

Luraghi, Silvia. 1998. The grammaticalization of the left sentence boundary in Hittite. The limits of grammaticalization, ed. by, A. Giacalone Ramat and P. Hopper, 189–210. Amsterdam and Philadelphia: Benjamins.

Wackernagel, Jacob. 1926–1928. Vorlesungen über Syntax, 2 vols. Basle: Birkhäuser.

———. 1953. Kleine Schriften, 2 vols. Göttingen: Vandenhoek & Ruprecht.

Watkins, Calvert. 1963. Preliminaries to a historical and comparative analysis of the Old Irish verb. Celtica 6. 194–249.

SILVIA LURAGHI

Warao

Warao is an Amerindian language spoken in the Orinoco Delta and adjoining areas in northeastern Venezuela and Guyana. Its speakers, the Warao, are the aboriginal population of the Delta, who have been present for as long as 9,000 years (W. Wilbert 1995:336). They have an estimated population of about 30,000, which is increasing. Although they form the second largest indigenous group in Venezuela, given that the total indigenous population of the country comprises only 1.5%, they are clearly a minority. The latest reliable official census rates 90% of the Venezuelan Warao as speaking their indigenous language and 48% of them as being bilingual in Warao and Spanish (Venezuela 1993). Especially in the central part of the Delta, missionary boarding schools have had the—albeit unintended—effect of language loss. There are therefore a growing number of monolingual Spanish speakers in this area and in the slums found in Tucupita and Barrancas.

The estimated 1,000 Warao on the Guyanese side are all in close contact with wider Guyanese society so that all of them can be assumed to be English speaking: some being bilingual Warao-English or even trilingual including Spanish. There is however the tendency to give up Warao (Forte 2000).

Social and Political Status of the Language

Up to now, Warao is a language with no genuine writing tradition. Nevertheless, recordings mainly of myths have been written down and published by anthropologists (J. Wilbert 1970; Heinen 1988), missionaries, and linguists (Osborn 1966a, b, 1967; Lavandero 1991, 1992). The diversified oral tradition includes different styles of speech and various genres. It differentiates for instance between '*denobo*' (<*deje nobo*: old stories) mythology on the one hand and '*deje jiro*' (new stories) or '*deje kwamotane abane*,'

which contain recent stories, gossip, and jokes, on the other hand. In addition, there is special language use during ritual as in shamanistic curing sessions and wailing for the dead (Briggs 1993).

The large majority of Warao speakers live in communities that are exclusively Warao. Outside these communities there is no social space for the usage of the indigenous language. In Guyana, there has been no bilingual education so far (Forte 2000). In Venezuela, indigenous languages were authorized for school education by a presidential decree in 1979, and official alphabets for most of them including Warao were released subsequently. Nevertheless, due to lack of political support these regulations were never put to practice on a larger scale (Villalón 1994). Warao language and culture are subject to strong pressure by the Spanish- and English-speaking majority, urging them to assimilate into the national Criollo or Creole culture of the respective countries. The awareness among the Warao concerning their language and culture as being endangered is just starting to rise, but still the future of their whole cultural heritage has to be regarded as most uncertain.

Dialects

There is no reliable information concerning the number of dialects. Some authors claim that there are none (Romero-Figeroa 1997); others have suggested that there are only slight regional variations (Osborn 1966a:108f.). For their part, Warao speakers claim not being able to understand speakers from certain other groups. Due to Delta internal migration, the situation has become regionally complex. Nevertheless, at least four major linguistic and cultural variants have been assumed by most researchers (Weisshar 1982; Wilbert 1957). Certainly, the different Warao groups are for ecological and historical reasons culturally quite heterogenous (Heinen and Garcia-Castro 2000).

Genealogy

Because of a lack of historical information, it is impossible to prove membership in a language family for Warao and it has therefore to be considered an isolate (Weisshar 1982). Certain resemblances with other languages have to be attributed to the phenomenon of a 'linguistic area', where languages of a certain region (in this case the Amazon and neighboring areas) show similarities across the boundaries of different language families.

Typology

Typologically speaking, Warao shows characteristics of an 'agglutinating' language. Languages of this type tend

to have one morpheme (smallest meaningful unit, which can be a word or a part of a word) per grammatical function. Larger words are then composed of a basic word or root and these morphemes. In the case of Warao, most of them—suffixes—occur after the root, although some of them—prefixes—are attached in front of it. For example, the phrase: '*ine najoro-turu-ae*', structurally 'I food+almost-would-have+completed-action', equals the English expression 'I almost ate.' (Please note that in these examples of Warao, the Spanish-based official spelling is used and the letter 'j' therefore corresponds to 'h' for English speakers. Hyphens are intended to clarify the structure of a word.)

Word Classes

The categories of adjective and noun are not clear-cut. A word can function as noun or adjective according to its place in a sentence and the suffixes it combines with. Nouns and adjectives do not have gender. There are furthermore 'noun–verbs' (Osborn 1967). Normally verbs and nouns can be distributionally differentiated according to the suffixes and prefixes (classed together as affixes) that can occur with them. Some roots, however, can combine with verb and noun affixes alike. To give an example, the root '*najoro*' (food) alone is a noun that can combine, for example, with the suffix '-*noko*' (place/instrument) to form '*najoro-noko*' (place of the food/instrument for eating). On the other hand, in the construction '*najoro-ya*' (literally, food+durative aspect), 'is/are eating', the same root functions as part of a verb combining with a verb suffix.

Tense, Mood, and Time

In Warao a wide range of such notions as tense, aspect, and mood are marked on the verb. The above-mentioned morphemes accomplish this task mainly suffixed to the basic verb form. In order to clarify this point, there is a nonexhaustive list of examples given below. However, as there is no consistent labeling in the literature for these forms, I will limit myself to give an approximate translation into English:

najoro-kitane (to eat)
i-najoro-kitane (to cause someone to eat/to feed someone)
najoro-kitia (going to eat immediately)
najoro-te (probably going to eat in the future)
najoro-ya (is/are eating at that moment/will surely be eating in the future)
najoro-ae (ate)
najoro-ya-ja (eating)
najoro-kuna (could eat)
najoro-turu-ae (was/were about to eat)

najoro-mejerei (in order to have him/her eat)
najoro-moana (has/have not yet eaten but should have done so)
najoro-komoni (cannot eat)
najoro! (eat!, singular)
najoro-kotu! (eat!, plural)
najoro-na-ja (is/are not eating)
najoro-na-tan-ae (has/have not eaten)
najoro-naka! (don't eat!, singular)
najoro-naka ja-kotu! (don't eat!, plural)

In this last example, '*jakitane*' (to be/to have) functions as an auxiliary verb taking on the imperative plural ending '*-kotu*' because it is impossible to attach it to the preceding suffix '*-naka*'.

Taken together, '*jakitane*' with '*tikitane*' (to say/to do) and '*takitane*' (to be/to do) form an interesting group. They can not only be used as auxiliaries and as suffixes but also as clause introducers. As in '*ta-kore*' (literally, this being so), then/when.

Analyzing '*najoro-na-ta-n-ae*' (has/have not eaten) from the above list, '*ta-n-ae*' really is the root of the verb '*ta-kitane*', followed by a marker for singularity of action '*-n-*', followed by the suffix expressing completeness of an action: 'food+not+do+singularity of action+completed action'.

Discourse Markers

In addition, there are morphemes that allow the modification of verbs, nouns, and adjectives and sometimes range over the whole sentence: So, for instance, '*-yama*' is a citation marker specifying that something is known from hearsay: *najoro-ae-yama* (it is said/someone told that he ate). It is prevalent in storytelling.

Other morphemes like '*-kore*' or '*-rone*', which mean 'while/if' and 'although', respectively, play the role of conjunctions connecting sentences or phrases: *nojoro-ya-kore dani mi-ae* (while I was eating I saw my mother); *najoro-ya-rone dani mi-ae* (although I was eating I saw my mother). For questioning a sentence, the suffix '*-ra*' is attached to the last word of that sentence.

Sentence Structure

As regards the 'basic' or unmarked word order in Warao, there is no general agreement. Some take a Subject–Object–Verb (SOV) word order to be basic (Osborn 1966b) while others argue in favor of Object–Subject–Verb (OSV) (Romero-Figeroa 1997). All that can be stated safely is that Warao is a 'verb-final' language.

In a number of languages, the copula 'to be' (as in an English sentence of the type: 'I am Warao') is not obligatory. The same holds true for Warao. Thus, '*ine warao*' (literally, I Warao) would be the equivalent of the English sentence.

The direct complements of a verb (subject, object) do not have case marking or gender and are often omitted, especially when they have been mentioned previously or are assumed to be known. Besides, there are stylistic reasons for omission. Nouns can take on suffixes (*-si/-ma/-to*) that specify that they are not the subject of a sentence: *dima-si konaria* (he takes father along) as opposed to *dima konaria* (father takes something/someone along).

A study of the 'pivot' in Warao is yet lacking. In languages with nominative and accusative case marking, the nominative (case of the subject of an intransitive and a transitive sentence) functions as the pivot. Thus, two sentences are connected having a common subject: 'the man ate an apple and smiled'. In ergative languages, however, which show the same case for the subject of an intransitive sentence and for the object of a transitive one, this juncture would be impossible. 'The man ate an apple and smelled good' would hence mean that the apple smelled good, not the man. In Warao storytelling, such connections are frequently made. This gives rise to the question of whether this language is syntactically ergative.

Person, Number, and Plurality of Action

Basically, there is no agreement for person and number on the verb, although interestingly singularity or plurality of the action itself can be marked and thus emphasized. Compare, for example, '*boro-te*' (will be jumping) with '*boro-bu-te*' (will be jumping repeatedly/a lot of people will be jumping). Here the morpheme '*-bu-*' underlines that the action is performed repeatedly either by one person or simultaneously by many.

The case of a morpheme that depicts singularity ('*-n-*') is more complicated as some verbs exist both with and without this morpheme and others only occur in one of these forms. An example of alternating verbs would be *ine oa-e* (I grabbed several things) in opposition to '*ine oa-n-ae*' (I grabbed once/one thing). It is not clear whether or not '*-n-*' is still a productive morpheme.

With respect to number marking on nouns, there is a suffix '*-tuma*' that is often regarded as plurality marker. But it is not obligatory and, especially when used with people, rather expresses the idea of 'the-ones- belonging-to' as in '*Maria-tuma*' (Maria and her friends/family).

Possession and Article

The possessive construction has the form 'possessor possession-marker+possessed-item': *dima a-janoko* (father's house, literally 'father his-house').

There is no definite or indefinite article in Warao. But person markers used in possessive constructions function like a definite article. In the following text recorded by the author, definiteness is assured in this way:

Tau-tuma manamo ja. Ta-kore a-raiba a-rajia tane.

Two womenfolk is/has. Then her older sister her younger sister so (literal translation).

There were two women: an older and a younger sister (free translation).

Future of the Language

Clearly a most interesting language with phenomena awaiting to be investigated, Warao presents a challenge to linguists. Unfortunately, like most amerindian languages the future vitality of Warao is not assured.

References

Briggs, Charles L. 1993. The pattering of variation in performance. American dialect research, ed. by Dennis R. Preston, 379–431. Amsterdam: Benjamins.

Forte, Janette. 2000. Amerindian languages of Guyana. As línguas amazonicas hoje, ed. by F. Queixalós and O. Renault-Lescure. Sao Paulo: Instituto Socioambiental.

Heinen, Dieter. 1988. Oko Warao: marshland people of the Orinoco-Delta. Münster: Lit.

Heinen, Dieter, and Alvaro Garcia-Castro. 2000. The multiethnic network of the lower Orinoco. Ethnohistory 47. 561–79.

Lavandero, Julio Perez (ed.) 1991. I Ajotejana, Mitos (Myths from Ajotejana). Caracas: Hermanos Capuchinos.

———. 1992. II Ajotejana, Relatos (stories from Ajotejana). Caracas: Hermanos Capuchinos.

Osborn, Henry. 1966a. Warao I: phonology and morphophonemics. IJAL 32.

———. 1966b. Warao II: Nouns, relationals, and demonstratives. IJAL 32.

———. 1967. Warao III: Verbs and suffixes. IJAL 33.

Romero-Figeroa, Andrés. 1997. A reference grammar of Warao. München, Newcastle: Lincom.

Vaquero, P. Antonio. 1965. Idioma Warao. Morfología, sintaxis, literatura (Warao language. Morphology, syntax, literature). Caracas: Estudios Venezolanos Indígenas.

Venezuela. 1993. Oficina Central de Estadística e Informática, Censo Indígena de Venezuela 1992, Tomo I (Venezuelan Indigenous Census of 1992, Part I). Caracas: OCEI.

Villalón, María E. 1994. Educación para indígenas en Venezuela: Una crítica razonada (Education for indigenous people in Venezuela: a reasoned critique). Caracas: CEVIAP.

Weisshar, Emmerich. 1982. Die Stellung des Warao und Yanomama in Beziehung zu den indigenen Sprachen Südamerikas nördlich des Amazonas: Studien zur genetischen und areal-typologischen Klassifikation (The position of Warao and Yanomama in relation to the indigenous languages of South America North of the Amazon: study in genetic and areal-typological classification). University of Tübingen: Dissertation.

Wilbert, Johannes. 1957. Prologo. Diccionario Guarao-Español, Español-Guarao, ed. by Sucre, 7–18. Basilio María de Barral. Caracas: Sociedad de Ciencias Naturales la Salle.

———. 1970. Folk literature of the Warao Indians. Narrative material and motive content. Los Angeles: Latin American Center University of California.

Wilbert, Werner. 1995. Conceptos etnoecológicos Warao. Naturaleza y Ecología Humana en el Neotrópoco/Nature and human ecology in the neotropics (Scientia Guaianæ. No. 5), ed. by Dieter Heinen, José J. San-José and Hortensia Caballero-Arias, 312–70. Caracas: Instituto Venezolano de Investigaciones Científicas.

<div align="right">STEFANIE HERRMANN</div>

Wayampi and Tupi-Guarani Languages

Wayampi is one of the more than 40 closely related languages that comprise the Tupi-Guarani language family. These languages descend from a common ancestor, referred to as Proto-Tupi-Guarani, whose speakers probably lived in or around the state of Rondonia in western Brazil. Long before the arrival of Europeans to the New World, extensive migration of small groups contributed toward the development of a large number of closely related languages, scattered throughout the Amazon region and beyond. As individual groups migrated, many of them through the Amazon rain forest, they eventually lost contact with each other. The language spoken by each group gradually changed in unique ways, until the speech of the individual groups was no longer mutually intelligible. There are no written records for this early history.

The first contacts that Europeans had with Tupi-Guarani languages were with Tupinambá and Guarani, at that time spoken along the eastern coast of South America. As explorers, anthropologists, and linguists came in contact with other indigenous groups whose languages showed a high degree of similarity in vocabulary and word structure to Tupinambá and Guarani, they referred to these languages as Tupi-Guarani languages.

The Tupi-Guarani family is one branch of a larger grouping of Tupi languages. Tupi-Guarani languages have a number of features that define them as a group, separate from the other Tupi languages. For example,

they have a set of free personal pronouns that are obvious cognates (having been derived from a single source); other Tupi languages do not have this set. Tupi-Guarani languages also share a set of verbal prefixes which refer to the subject: *a-* 'I', *oro-* 'we exclusive (he/she and I)', *ja-* 'we inclusive (you and I)', *ere-* 'you (sg)', *pe-* 'you (pl)', and *o-* 'he/she/it/they'. They also share a number of basic vocabulary items that are different from forms in other Tupi languages.

Wayampi

Wayampi (with variant spellings Guayapi, Waiãpi, Wayãpi, Wajapĩ, Oiampi, Oyampi, and Oyampik) is the autodenomination of a language group of approximately 1,800 speakers, living on the two sides of the Brazil–French Guiana border. They are one of the three Tupi-Guarani language groups residing on the north side of the Amazon River.

Originally residing near the mouth of the lower Xingu River, the Wayampi migrated north across the Amazon and up the Jari River in the mid-1700s. During the first half of the twentieth century, there were four dialect groups of Wayampi: one on the Oiapoque and Camopi rivers in French Guiana, another on the Cuc River in the Brazilian state of Amapá, a third group living east of the Cuc River in the area that they call *Kumakary [kumakarɨ]*, and a fourth group living farther south, on small tributaries of the Amapari and Nipuku rivers.

The group on the Cuc River had ongoing contact with the groups in both Oiapoque and Kumakary. Around the middle of the twentieth century, the Amapari group came in contact with the group from Kumakary, and the two groups eventually merged. Through these contacts the people from the Amapari area learned of the group on the Cuc River, with whom they began to develop a trading relationship. However, this relationship was interrupted around 1970 when a dysentery epidemic in the Cuc River area took the lives of a large number of people. After that, most of the survivors moved to French Guiana, where they merged with the Oiapoque group.

In spite of their relative isolation from the outside world, the Wayampi were visited by French explorers in the late nineteenth century, among them Henri Coudreau (1892), who published an impressive word list.

The general phonological pattern for Wayampi is consonants alternating with vowels. The consonant set includes *p,t,k,kʷ,ʔ,m,n,ŋ, ŋʷ,r,s,h*, and, in the Amapari dialect, *β*. There are two semivowels, *w* and *j*. The set of vowels includes *i,e,ɨ,a,u,o.* There is an oral set and a nasalized set of vowels. When nasalization occurs, it affects the entire morpheme and not just a single vowel, such as *apasĩ [ãpãsĩ]* '(to) tie'. In the Amapari

dialect, there is only one back nasalized vowel, and younger speakers are merging the central nasalized vowels as well.

Grammatically, Wayampi has a number of features that are different from Indo-European languages but are quite common to Tupi-Guarani languages and, in fact, to Amazonian languages in general. Like other Tupi-Guarani languages, Wayampi is agglutinative, which means that words can take multiple prefixes and suffixes. This is most visible in verbs. In the following example from the Cuc dialect, the verb stem (*kusu* 'wash') is preceded by an incorporated object (*po* 'hand'), which is in turn preceded by three prefixes. There are also two suffixes following the verb stem:

n-o-i-po-kusu-ɨtɨ-i
not-he-self-hand-wash-never-not
'He never washes his (own) hands.' (literally, 'he never hand-washes himself')

Typologically, one of the areas of greatest interest to linguists studying Wayampi and related languages is the system of person markers which occur on verbs. Transitive verbs, those which express an action on an object, are prefixed by person markers from one of two sets, one referring to the subject and the other referring to the object. For example, there are two prefixes referring to first-person singular: *a-* 'I' (subject) and *e-* 'me' (object), as in the following examples: *a-pota* 'I like (him)' and *e-pota* '(he) likes me'. There are complicated rules governing the choice of sets, but, in general, first person ('I/we'), whether subject or object, is considered the most important and must be referred to on the verb, as in the examples above. Second person ('you') is second in priority, and third person ('he/she/it/they') has the lowest priority. This is one type of 'split-ergative' system. The 'subject' prefix is also used for intransitive verbs, and the 'object' prefix is used for a set of stative verbs: *a-poreŋeta* 'I speak' and *e-karaɨ* 'I have a fever'. This type of system is characteristic of Tupi-Guarani languages. Variations of split ergativity occur in other Amazonian languages as well.

When two verbs are perceived by the speaker to be part of a single action, with the same subject, they occur together in a single phrase as 'serial verbs'. If the second verb is transitive, as in the second sentence below, it must take an object prefix rather than a subject prefix. (In third person, there is no grammatical distinction between masculine, feminine, and neuter.)

o-o   o-jau
he-go he-bathe
'He went to bathe.'

o-nupã   pira i-juka
he(subj)-beat fish it(obj)-kill
'He beat the fish and killed it.'

Unlike other Tupi-Guarani languages, Wayampi does not take a special suffix to mark the second serial verb.

Noun phrases are short and are generally limited to one other constituent besides the noun. They follow one of the following patterns:

Demonstrative-Noun:
aʔe kuimaʔe   'that man'
that man
Number-Noun:
moapɨ kuimaʔe   'three men'
three   man
Noun-Relative Clause:
kuimaʔe i-tekorã maʔẽ   '(the) man that is sick'
man   him-sick RELATIVIZER
Genitive-Noun:
papa po   'father's hand'
father hand
Noun-Postposition:
ɨa   pupe   'in the canoe'
canoe in

Pronominal possessors are indicated by the 'object' prefixes: *e-po* 'my hand', *ne-po* 'your (sg) hand'. The same is true for the pronominal object of a postposition: *e-pupe* 'inside me', *i-pupe* 'inside it'.

Wayampi is what is referred to as a 'flexible word-order' language. Unlike English, which generally puts the subject before the verb and the object after the verb, Wayampi uses other criteria to determine what information goes before the verb and what goes after. For example, new information is introduced before the verb, but when it is referred to again, it occurs after the verb. In the following example, the participants of an all-night festival announce their intentions:

jeʔiwe   ture oro-itɨ-ta,
early.morning flute we(exclusive)-quit-future

eʔi kupa
said plural.subject

''Early in the morning we will quit playing flutes', they said.'

Both the time and the object (flute) are new information and come before the verb (underlined). Once introduced, they are subsequently referred in postverbal position, as in the following example:

o-itɨ   jeʔiwe   ture kupa
they-quit early.morning flute plural.subject
'They quit playing the flutes early in the morning.'

In spite of the flexible word order, there is no grammatical case marking on nouns to distinguish subjects from objects. Often only one noun is actually specified in the sentence and the other is indicated only by the prefixing on the verb or is simply understood from context. In the sentences above, it is obvious that *ture* 'flute' is an object, not a subject, since the verb (*itɨ* 'quit') requires an animate subject.

Very little has been published on the discourse of other Tupi-Guarani languages, so it is not possible to say to what degree Wayampi discourse is characteristic of the family as a whole. However, this type of discourse system is not unusual among American indigenous languages.

Other Tupi-Guarani Languages

The Tupi-Guarani family and member languages have been studied extensively from the perspectives of historical linguistics and syntactic typology. Reconstructions have been done of Proto-Tupi-Guarani grammar and sound system. Using the criteria of a shared history of phonological changes from the reconstructed proto-language, Aryon Rodrigues has divided the Tupi-Guarani languages into eight subgroups.

Languages in subgroup 1 are Guaranian languages, spoken in southern Brazil, Paraguay, Bolivia, and northern Argentina. One form of Guarani, now referred to as Old Guarani, has been known from the time of the Spanish explorers who first came to the area south of São Paulo. This language was documented and described by the Jesuit priest Antonio Ruiz de Montoya (1639). In time, Guarani came to be used extensively among colonists. Gradually, a no indigenous variety of the language developed, which is now spoken by 95% of the population of Paraguay.

Languages in subgroup 2 are spoken in Bolivia: Sirionó and Guarayu. Languages in subgroup 3 were spoken along the Atlantic coast and up the Amazon river. The most well-known language from this group is Tupinambá, which was documented extensively by another Jesuit priest, Joseph de Anchieta (1595). Although now extinct, Tupinambá was used extensively for many years among colonists. Finally, a royal decree in the eighteenth century dictated that Portuguese citizens use Portuguese rather than an indigenous language as their means of communication. Even so, the nonindigenous variety of Tupinambá, now called *Nheengatu*, is still spoken today as the mother tongue of a nonindigenous population of about 3,000, in the upper Amazon region. Brazilian Portuguese has a large number of words borrowed from Tupinambá, especially place names and terms for flora and fauna native to Brazil. Even English has borrowed words from Tupinambá: jaguar, toucan, piranha, tapioca, and manioc.

Speakers of languages of subgroup 4 reside at the eastern edge of the Amazon rain forest and beyond, in

the Brazilian states of Tocantins, Maranhão, and Pará. They include Asuriní of Tocantins, Parakanã, Tapirapé, Guajajára.

Languages of subgroup 5 are west of subgroup 4, in Pará and northern Mato Grosso. They include Asuriní of Xingu and Kayabí.

Still further west, in Amazonas, Rondônia, and Acre, are the speakers of subgroup 6: particularly the Kawahíb dialect cluster. These groups live closest to the area where their Proto-Tupi-Guarani ancestors probably lived.

Subgroup 7 includes just one language: Kamayurá, located in the Xingu Park in Mato Grosso.

Subgroup 8, of which Wayampi is a member, includes the three language groups located on the north side of the Amazon, who live in southern French Guiana or northern Brazil. It also includes several language groups who live on the south side of the Amazon in the states of Pará or Maranhão, among them Guajá and Urubú-Kaapor.

References

Anchieta, Joseph de. 1595. Arte de gramatica da língua mais usada na costa do Brasil. Coimbra: Antonio de Mariz.

Coudreau, Henri. 1892. Vocabulaires méthodiques des langues Ouyana, Aparai, Oyampi, Emerillon. Paris: Bibliothèque Linguistique Américaine.

Derbyshire, Desmond C., and Geoffrey K. Pullum. 1986 and 1998. Handbook of Amazonian languages, vols. 1 and 4. Berlin: Mouton de Gruyter.

Jensen, Allen. 1994. Wayampi. Typological studies in negation, ed. by Peter Kahrel and René van den Berg. Amsterdam/Philadelphia: John Benjamins.

Jensen, Cheryl. 1999. Tupi-Guarani. The Amazonian languages, ed. by R.M.W. Dixon and Alexandra Y. Aikhenvald. Cambridge: Cambridge University Press.

Navarro, Eduardo de Almeida. 1998. Método Moderno de Tupi Antigo: a língua do Brasil dos primeiros séculos. Petrópolis: Editora Vozes.

Rodrigues, Aryon Dall'Igna. 1986. Línguas brasileiras: Para o conhecimento das línguas indígenas. São Paulo: Edições Loyola.

Ruiz de Montoya, Antonio. 1639. Tesoro de la lengua Guaraní. Madrid: Juan Sanchez.

CHERYL JENSEN

Weinreich, Uriel

An authority in a number of disciplines, Uriel Weinreich made significant contributions to several aspects of linguistics and was particularly interested in those facets dealing with the interaction of language and society. His specialties included but were not limited to language change, semantic theory, and the Yiddish language. Born in Lithuania, Weinreich moved to the United States in his early teens, served in the US Army during World War II, was naturalized as an American citizen soon after, and worked for a brief time with the State Department. As a scholar, Weinreich made his home at Columbia University, where he received his undergraduate and graduate degrees and later taught linguistics and Yiddish until his death at the age of 40.

One of Weinreich's major contributions to linguistics was his work on language change and variation. Under the guidance of his advisor Andre Martinet, Weinreich wrote both his master's thesis and his dissertation on the role of language contact in linguistic change, and this work later served as the foundation for his book *Languages in contact* (1953). In his article 'Is a structural dialectology possible?' (1954), Weinreich called for scholars working in the seemingly adverse frameworks of structural linguistics and dialectology to find a common ground on which to work together toward a better understanding of language. Weinreich also worked on a model of language change that would be completed and published posthumously by his students William Labov and Marvin Herzog in an article entitled 'Empirical foundations for a theory of language change' (1967). In this work, Weinreich et al. argue that a model of language that 'accommodates the facts of variable usage and its social and stylistic determinants not only leads to more adequate descriptions of linguistic competence, but also naturally yields a theory of language change that bypasses the fruitless paradoxes with which historical linguistics has been struggling for over a half century' (Weinreich, Labov and Herzog, 'Empirical foundations for a theory of language change', 1967:99).

As part of their proposal, Weinreich et al. downplay the role of the idiolect in linguistic change in favor of an explanation that takes into account the interrelationship of linguistic and social factors. In this respect, Weinreich's ideas had a profound effect on Labov, for whom Weinreich served as thesis and dissertation director, in his pioneering work in the emerging field of sociolinguistics.

Weinreich also made a significant contribution to the study of semantics. Frustrated with a doctrine of

linguistics that chose to ignore the issue of semantics because of the difficulty it created, Weinreich saw promise in the generative grammar espoused by Chomsky (1957, 1965) because he believed that this formal approach could be used to account for meaning, despite Chomsky's avoidance of semantics in his own work. Weinreich was, however, critical of many of the approaches to semantics that were in vogue during his career. For instance, he asserted that 'systematic semantics is doomed to an isolated place within linguistics as a whole so long as paradigmatic relations are the only form of patterning to which attention is given. It is undoubtedly important to understand how the meaning of a word in a vocabulary is determined by other words in the same vocabulary; however, an account must still be given of the way in which the meaning of a sentence is composed out of the meanings of individual words' (Weinreich, *Explorations in semantic theory*, 1972:112).

Weinreich also took exception to the theoretical assumption of many working in semantics that semantics begins where syntax ends; instead, Weinreich believed that syntax and semantics are so intertwined that only a theory taking into account the interrelation between them could adequately represent the complexity of language. He pointed out that although generativists postulated syntax and semantics to be separate domains, they had never succeeded in discovering the boundary between the two. Weinreich was particularly critical of the Katz–Fodor model of generative semantics (see Weinreich 1972) on the grounds that their analysis focused on a limited part of semantic competence and concentrated too much on ambiguity. Weinreich argued that 'grammatical theory is not required to explain how a hearer of such ambiguous expressions guesses which of two deep structures is represented by a given occurrence of a surface structure, nor is the goal of grammatical theory limited to the calculation of such ambiguities' (Weinreich, *Explorations in semantic theory*, 1972:18–19). Weinreich was also critical of the Katz–Fodor model because he said that it ignored semantic deviations, instead of trying to account for them.

Like his father, Yiddish scholar Max Weinreich—a driving force behind the YIVO Institute for Jewish Research and a proponent of the claim that Yiddish was a unique creation of Jews and not simply derived from German—Uriel Weinreich was a scholar of Yiddish and worked hard for the preservation of the language, collecting a great number of Yiddishisms for various books on the subject. His textbook on the subject, *College Yiddish*, was first published in 1949 and has since been published in numerous editions. In 1959, Weinreich and his wife, the folklorist Beatrice Silverman Weinreich, began interviewing Jews living in Israel, France, the United States, and Canada, with a questionnaire comprising more than 3,000 questions and taking some 15 hours to complete. This work served as the foundation for a comparative study of Jewish culture throughout Europe before World War II that culminated in the *Language and cultural atlas of Ashkenazic Jewry*, which Uriel directed until just before his death. Now under the direction of Herzog, the first volume of the project was published in 1991. The *Modern English–Yiddish and Yiddish–English dictionary*, which Weinreich compiled by applying his theory of systematic lexicography, was published posthumously in 1968. To honor the man and his work, the Uriel Weinreich Program in Yiddish Language, Literature, and Culture was established as a summer program in 1968, cosponsored by Columbia University and the YIVO Institute for Jewish Research.

Biography

Uriel Weinreich was born in 1926 in Vilnius (Russian *Vilna*), Lithuania. He moved to the United States with his family in 1940, when his father, Yiddish scholar Max Weinreich, joined the faculty of New York City College. He served as first lieutenant in the US Army from 1944 to 1946 and was naturalized as an American citizen in 1945. After his term of service, Weinreich enrolled in Columbia University and received his BA in 1948, his MA in 1949, and his PhD in 1951, completing a thesis and dissertation under Andre Martinet, work which would eventually culminate in his book *Languages in contact* (1953). He performed fieldwork in Switzerland from 1949 to 1950 under a research fellowship from the American Council of Learned Societies. He worked as an editor and information specialist with the State Department from 1951 to 1952, and was Chair of the Department of Linguistics at Columbia University. Weinreich became the first Atran Professor of Yiddish Language, Literature and Culture at Columbia University in 1952, and was Member, Linguistic Society of America in 1951, serving as vice president in 1964, and co-editor of Word from 1953 to 1960. He began work on the *Language and cultural atlas of Ashkenazic Jewry* in 1959 and served as its director until his demise. Weinreich continued teaching until two days before cancer took his life on March 30, 1967.

References

Chomsky, Noam. 1957. Syntactic structures. The Hague: Mouton.

———. 1965. Aspects of the theory of syntax. Cambridge: MIT Press.

Labov, William. 1994. Principles of linguistic change, Vol. 1: internal factors. Oxford: Blackwell.

Malkiel, Yakov. 1967. Uriel Weinreich (1926–1967). Language 43. 605–7.

Weinreich, Uriel. 1949. College Yiddish: an introduction to the Yiddish language and to Jewish life and culture. New York: YIVO Institute for Jewish Research.

———. 1953. Languages in contact: findings and problems. New York: Linguistic Circle of New York.

———. 1954. Is a structural dialectology possible?. Word 10. 388–400.

———. 1968. Modern English-Yiddish, Yiddish-English dictionary. New York: YIVO Institute for Jewish Research.

———. 1972. Explorations in semantic theory. The Hague: Mouton.

———. 1980. On semantics, ed. by William Labov and Beatrice S. Weinreich. Philadelphia, PA: University of Pennsylvania Press.

Weinreich Uriel, William Labov, and Marvin I. Herzog. 1963. On the semantic structure of language. Universals of language, ed. by Joseph H. Greenberg. Cambridge, MA: MIT Press. Empirical foundations for a theory of language change. Directions for historical linguistics, ed. by Winfred P. Lehmann and Yakov Malkiel. Austin: University of Texas Press. 1968.

LAMONT ANTIEAU

See also **Labov, William; Martinet, Andrè**

Western Caribbean Creole(s)

Western Caribbean Creole (WCC, hereafter) belongs to the western branch of Atlantic, English-based Creoles. It has features common to the broader linguistic designation, West Indian Creole (Winford 1985; Grimes 2001; Roberts 1996). Varieties of WCC developed in predominantly English-speaking western Caribbean nations, including Jamaica, Belize, and Barbados. Believed by some linguists (Hancock 1980; Niles, 1980) to be a dialect of English, Barbados Creole, popularly known as *Bajan* or 'Barbadian nonstandard English' (Fields 1995), 'has the fewest Creole features' (Roberts 1996:87), and its Creole status is somewhat controversial. It differs from other varieties of West Indian Creole primarily phonologically (Grimes 2001). Researchers such as Cassidy (1980) and Fields (1995) maintain that early Bajan was a Creole. Fields (1995) makes a case for its Creole ancestry, arguing that contemporary Bajan is simply a Creole-in-transition. While Hancock (1980) and Niles (1980) agree that early Bajan was probably a Creole, they differ in their accounts of precisely when Barbados Creole existed. Cassidy (1986) contends that Barbados Creole could only have developed after 1650, when conditions were favorable for its genesis.

The largest number of speakers of WCC is found in Jamaica, followed by Belize. In Jamaica, this language is known as Jamaican, Jamaican Creole (English), or Bongo Talk [*Bongo* is an Afro-Caribbean dance]. Linguists tend to use the label West(ern) Caribbean Creole. The variety spoken in Belize—slightly different—is referred to as Creola, Kriol, or Belize Creole (Boland 1986, 1991; Escure 1981, 1997; Greene 1999). So frequent is the use of WCC in these countries that it is the primary language and preferred lingua franca of the majority of residents. In Jamaica, for instance, the official language is English, but the use of WCC is more widespread. It, therefore, functions like the quasi-official language.

Despite its frequent use, WCC continues to be stigmatized, particularly by its educated speakers (Le Page 1966; Nettleford 1966; Allsopp 1996; Spears and Winford 1997), many of whom use this code as well. WCC may have started out as a pidgin or, in Romaine's (1994) opinion, as a *jargon* or a short-lived pidgin. Instances of the earlier form are found in anglophone Caribbean work-songs, and in dancing and folk songs from 1793 to 1907 such as the anonymous 'Freedom a Come Oh!' (Freedom, Here I Come!, where the 'Oh' might very well be a remnant of the Kwa honorific particle 'o(h)' used extensively in Yoruba and Nigerian Pidgin), 'Negro Song at Cornwall', and 'Sangaree Kill de Captain' (Sangaree Killed the Captain) (see Burnett 1986:3–12). Sprinklings of WCC can also be found in veteran Jamaican writer, Roger Mais's (1905–1954) *The hills were joyful together*, and in Vic S. Reid's works. WCC is presently the native or first language of a large percentage of users; hence the descriptive term, *Creole*.

Creolization is the term used to describe the formative process it is believed to have undergone. WCC and its eastern Caribbean and West African equivalents, such as Guyanese (see Bickerton 1973; Rickford 1987), Krio, or Nigerian Pidgin (Faraclas 1996), respectively, exhibit a *Creole continuum*, or a range of varieties. At one end is the *basilect*, a variety with minimal structure and associated functions, and at the other end lies the *acrolect*, a fully expanded variety, used, like any primary language, for all possible language

functions. The *mesolect* constitutes the mid-range. It is more compact than the acrolect, but serves a greater range of functions than the basilect. The different lects tend to be socially stratified (Pandey 1997). Mesolectal WCC is a combination of English and Creole words. The acrolect tends to contain few Creole words. Examples include:

(1) Acrolect: 'Don't lie to yuself', 'Dis is crazy' (Patterson 1964:16, 186).
(2) Mesolect: 'No matter if he is Jesus or him is Jah. Him not gwan like dis one lickle bit' (Cliff 1996:16).

[Translation: It doesn't matter if he is Jesus or Jah (Haile Selassie). He will not like this even a little.]

(3) Basilect: 'God nuh mus' be Hinglish. But me did' 'ear once dat Jesus did' ave bad 'air. Mus' be one joke 'pon we' (Cliff 1996:17–18).

[Translation: It seems God must be English then. But once I heard that Jesus had bad hair (reference to him as African). It must be a joke being played on us.]

The speech of the poor and less educated Barbadian is very similar to the mesolect found in neighboring anglophone nations, leading linguists like Alleyne (1979) and Fields (1995) to contend that it is in the mesolectal stages.

The term creole continuum was first used by creolist Derek Bickerton (1981), and has been applied to a variety of Pacific and Atlantic Creole contexts (colonized coastal territories). It is also termed a post-creole continuum (Rickford 1987), as it mirrors a process of decreolization or progressive assimilation toward a standard language. Bajan might be one such example (Cassidy 1980, 1986; Rickford 1992; Rickford and Handler 1994; Fields 1995). In Cassidy's (1980) view, Bajan decreolized earlier than other Caribbean Creoles. Decreolization is not necessarily chronological. In other words, it is more often the case that the different lects coexist, rather than in isolation, changing over time in lectal designation.

Jamaican Creole is viewed as a paradigm example of WCC (Lalla and D'Costa 1990; Mufwene 1993:142). Many Jamaicans (Bennett 1966, 1981; MacCallum 1999; Donnell and Welsh 1996) refer to it as *Patois*, the French term for a pidgin. The presence of large numbers of French migrants, and the island's proximity to French-colonized Caribbean nations might have something to do with this label. The anglicized Patwa is also used by some. Like all varieties of WCC, Jamaican Creole is English-based, which means that the bulk of its vocabulary is drawn from English, while its grammatical and discourse structure mirror the influence of other (non-European) languages. It contains ingredients from the different colonizing and Diaspora languages

that constitute the sociolinguistic mosaic of this region (Burnett 1986:xxv). In the words of well-known Jamaican poet and storyteller, Louise Bennett:

> The Jamaican patois has obviously Standard English as its roots but has been spiced with the tongues of the Arawak and Carib Indians (the original inhabitants; Africans brought over as slaves; Spaniards, the first conquerors; the migrating French, Chinese and East Indians; the Colonizing British; and in recent times, visitors from North America and the Rastafarian religion (1981:i).

In many ways, then, it is a linguistic chutney of sorts, much like chutney music, a variety of Soka music composed in Trinidad. Bennett draws heavily on WCC in her works. Exemplary poems include, among others, 'Jamaica Patois', 'Jamaica Obeah' (Jamaican Juju Man), 'Weh Dem Deh?' (Where are They?), 'Solja Bwoys' (Soldier Boys), 'White Pickney', 'Show Yuh Foot' (Show Your Foot), 'Him de Yah' (He is Jesus), 'Me Bredda' (My Brother), 'New Govanah', 'Train Leff Miss Hayes', 'Pass fe White', and 'De Royal Commotion'.

One of the most widely studied Caribbean Creoles, the variety used in Jamaica has the longest-standing history of use in Caribbean literature, including folklore (see McKay 1933; Bennett 1966; Burnett 1986; Mordecai and Morris 1980; Senior 1989; Cliff 1996; Donnell and Welsh 1996; Patterson 1998; MacCallum 1999, among others), music, and other artistic media. The international attention and recognition Jamaican music, dub poetry and music (see Allen 1990; Breeze 1991, 1997; Smith 1979), and Rastafarianism have attracted have helped spread WCC to other parts of the Caribbean, and have helped familiarize others outside the region with its unique rhythms, sounds, and structure. Much like rap is to the African American community in the United States, dub poetry and music are signature forms of Black folk-culture in Jamaica and the Caribbean as a whole.

Dub poetry refers to 'Lines meant to be spoken, generally to a two-beat rhythm, and dealing mostly with the life experiences and/or point of view of Black people in or from the Caribbean' (Allsopp 206). Dub music, derived from Reggae, is also a two-beat rhythm created principally by bass and drums. Both art forms employ principally WCC lyrics. When one speaks of Creole, it is no wonder, then, that the layperson generally thinks of Jamaica(n). The increasing use of 'I', a symbol of self-respect and group solidarity, in the Jamaican variety of WCC (see Pollard 1986), as a popular substitute for the Creole 'mi', and in words like 'I-lect' (Rasta dialect or Rasta Talk), 'I-quality' (equality), 'Idren' (brethen), and 'I're' (alright) can be attributed to Rastafarianism, and are a reminder of the Rastas's reverence for Ras Tafari (Emperor Haile Selassie).

Exemplary proverbs in WCC include 'Cockroach no business in fowl-yard' (i.e. Mind your own business), 'De higher monkey clim, de mo im expose' (i.e. promotion has its price—pretentiousness), and 'Yu lay wid dawg, yu get wid flea' (i.e. bad company corrupts). WCC is so important a part of Belizean and Jamaican culture that visitors to these countries are advised to familiarize themselves with the distinctive features of WCC (see Porter and Prince 2000; Brosnahan and Keller 1997). As regards Jamaica, Porter and Prince (2000) offer the following overview and advice:

> The unofficial language is a patois.... This archaic and simplified structure, coupled with African accents and special intonation, can make the language difficult to understand. Some of the most interesting anecdotes and fables are usually told in the patois, so understanding its structure can add to your insight into Jamaican culture.... Proverbs and place-names [e.g. Red Gal Ring] express some of the vitality [of patois].... The patois has been embellished and altered with the growth of Rastafarianism. (196)

In their focus on Belize, Brosnahan and Keller (1997) describe WCC as Belize's 'own colorful Creole dialect' (276). Bolland (1977) describes it as 'a means of expression through which proverbs, sayings, and folktales' convey 'African values and wisdom' (12), and as a code language used to satirize the British.

In Belize, the Kriol Projek, a national organization, continues to play a seminal role in educating citizens about the culture and creativity inherent in Kriol. This organization, and the present Governor General have compiled and printed folktales in Kriol for use in (Belize's) schools. As few Belizeans are used to seeing Kriol in writing, these publications are accompanied by CDs, so that students can hear the Kriol.

The presence of branches of the University of the West Indies in both Barbados and Jamaica helps ensure that WCC continues to be well researched. Some of the earliest studies of WCC include F.G. Cassidy's 1961 text; B.L. Bailey's (1962) *Language guide to Jamaica*, for US Peace Corp volunteers, her (1953) M.A. thesis, a version of which was published in 1966; and Cassidy and Le Page (1967). In the United Kingdom, WCC is the most identifiable and influential of all Caribbean languages owing to the large numbers of Jamaican immigrants. In fact, the evolution of British Black English has been attributed, in large part, to the frequency of use of Jamaican Creole in the United Kingdom (Sutcliffe and Wong 1986). WCC is very similar to West African Pidgin English, leading many creolists (Le Page 1966; Dalphinis 1985; Faraclas 1996; Roberts 1996) to argue in favor of the Afrogenesis hypothesis, which traces its roots to Africa, and highlights the influence of West

African languages. According to one view, for instance, 'Created in West Africa, ... work pidgins were transported across oceans where they took on new roles as lingua francas among the enslaved....' (McWhorter 1997: 240). Dalphinis (1985) observes that '[B]oth in the analysis of the Creole languages of the Caribbean as well as Caribbean culture, it is clear that Europe has affected Africa but equally Africa has affected Europe; the words are European, but the syntax is African' (95).

Representative features of WCC include (see Winford 1985, 1993; Roy 1986; Escure 1978, 1983; Patrick 1993; LaCharité 1999, among others):

(1) The zero copula or 'be' verb before adjectives, as in 'Di pikni sik' (The child is sick) or 'Di chile/chilin sik' and 'im hongry' (She or he's hungry) in Kriol, 'im [Bradda Anancy] always foolin every oda creature' (MacCallum 1999:19), and 'We making history' (Cliff 1996:5).

(2) Nonmarking of the suffix in the third-person singular present, as in 'I laik it' (Kriol for 'She or He likes it'), and the use of nonfinite verbs to convey multiple meanings, e.g. 'Dat sopoz to daan' (Kriol for 'That's supposed to have been done') and 'De chilin jos ron' (Kriol for 'The children just keep running around').

(3) Spelling pronunciations; words are pronounced the way they are spelled, and vice versa, e.g. 'dat' for 'that', 'de' for 'the', 'dis' for 'this', and the use of 'Dem' for many English words, including plural marker, as in (15); and 'they', 'them', and 'their(s)' as in 'Mi a sel dem az mi laik' (Bailey 1966:57), meaning 'I am selling them as I like'.

(4) Nonmarking of possessive case; possession clarified through noun juxtaposition, as in 'so many people life' (MacCallum 1999:20), 'she sister husban get one job up a Mona' (from Louise Bennett's 'Jamaica Patois'), 'mai li breda fren' (i.e. Kriol for 'My little brother's friend'), and 'person business' in (5).

(5) Multiple meanings associated with 'for', pronounced 'fi', as in 'Dem always start wid anancy ... lookin fi smbady... im always wan fi push up imself inna every oda person business' (MacCallum 1999:19). [Gloss: They always start with Anancy looking *for* somebody.... He always tries *to* stick his nose in other people's business.] An example from Bailey (1966:127) is 'Im kom ya fo bada mi', meaning 'He has come here to bother me'.

(6) Multiple meanings for 'no' (i.e. as 'not', 'don't', 'doesn't', and 'didn't'); double negatives

and preverbal negation, e.g. 'Im no know what mek anancy stay so wutless... and no seem fi care' (MacCallum 1996:19) [He didn't know what makes/made anancy so worthless... and not seem to care].

(7) Absence of auxiliary verbs, including 'do' in negations, as in 'I no kom bak' (Escure 1980:34), meaning 'She or he didn't come back', 'Me no 'ave not'in fe hide' (Patterson 1964:62), meaning 'I don't have anything to hide', and 'People talking around you' (V.S. Reid, 'New Day', p. 198), as well as the use of the double negative, as in Black English Vernacular.

(8) Zero tense specifiers or the default use of the present tense to represent all tenses. Tense is clarified through the conjoining of base verbs and/or context, e.g. (7), 'Toder mornin me go ask her wat she tink about de war' (from 'Jamaica Patois'), and 'Licklemos' she draw me in. But me stop because of lickle monies...' (Cliff 1996:19), which translates as 'A little more and she draws/drew me in. But I stopped because I didn't have a lot of money.'

(9) Nonmarking of plurals, as in 'Tu Bwai' (Kriol) or 'Tu bwoy' (Jamaican Creole), and 'Ho moch man?' (Kriol for 'How many men?')

(10) The use of 'vex' to mean 'distraught', 'troubled', 'disturbed', and 'disoriented', as in 'anancy being anancy was vex' (MacCallum 1999:20). Given the pronunciation differences between Kriol and Jamaican Creole, this word is prounced and written as 'bex' in Kriol. This probably has to do with the Spanish influence in Belize; /b/ and /v/ tend to be conflated in Spanish. Hence, 'I bex' would mean 'She or He's mad/frustrated.'

(11) The use of 'tief' as a verb for 'steal', 'stole', and 'stolen', as in 'afta anancy tief away dat magic calabash' (MacCallum 1999:19) and 'Tiefin de cow' (MacCallum 1999:22). An example from Kriol is 'Bra Anansi de plan tu go tiif ya hag' (Escure 1981:34) or 'Brother Anansi is planning to go steal a hog.'

(12) A relatively compact vocabulary (in contrast to Standard English) in which many words mean more than one thing, depending on the context, e.g. 'me' for 'me', 'my', and 'mine;' and *man hag* and *uman hag* versus the distinct English words, *boar* and *sow*.

(13) Related to (12), two or more meanings are assigned to a single verb; for instance, 'get' expresses both 'possession' and 'existence', as in 'Im get wan pikni' (She or He has one child) and 'im get good life.'

(14) Consonant reduction, as in 'im' vs. 'him', 'an' for 'and', 'a' for 'at', 'ow' vs. 'how', 'wha' vs. 'what', 'round' for 'around', 'ac' vs. 'act', and 'lookin' vs. 'looking'. This is sometimes accompanied by the prevocalic insertion of sounds like aspiration or /h/. Examples from Cliff (1996) include: 'Is hokay wid me, wunna know. Dem nuh 'tink like brute—hall a dem?' (10) [Translation: It's okay with me, you (plural) know. Don't they think like brutes—all of them?]

(15) 'Cyaan' or 'Kyaan' has context-dependent, opposite meanings, including 'can' and 'can't', as in 'di biebi-dem kyaan fiid themselves' (Bailey 1966:132), meaning 'The babies can't feed themselves' and 'Lord Jesus, hear me nuh, why dem cyaan behave demselves' (Cliff 1996:18), meaning 'Lord Jesus, please hear me. Why can't they behave themselves?' One of dub poet Michael Smith's poems is 'Mi cyaan believe it.'

(16) Emphatic 'nuh' as in 'God nuh say so?' (Cliff 1996:18), meaning 'Didn't God say so?'; Africanisms like *deh* and *fe*; and other culture-specific terms such as *Myal-man, Obeah-man*, and *Jumby-man*; e.g. 'Hear me, Myal-man. Hear me, Obeah-man. Hear me, Jumby-man.... Suffering nuh mus' be meant for we.... How long mus' we wait to get t'rough?' (Cliff 1996:16–17), which is essentially an invocation that translates as 'Please hear me Myal-man (one who performs good miracles). Please hear me Obeah-man (one who works in the 'dark' arts). Please hear me Jumby-man (one who is like a spirit or ghost). Suffering must be meant for us. How long must we wait to get through?'

Most of these are characteristic of West African Pidgin. WCC is an oral tongue, so it is written in accordance with a writer's perception of how it sounds. It is believed to echo the orality of African folklore and traditions (see Burnett 1986; Bennett 1940, 1966, 1979; Toolan 1992; Brathwaite 1993; Lalla 1996; Warner-Lewis 1997; Balutansky and Sourieau 1998). Those who employ it argue in favor of its legitimacy. Like Jamaican poet Claude McKay, who was the first to employ solely WCC in his first two (1912) books, one of the earliest Caribbean writers to make use of WCC in her writing, despite the criticism she received for doing so (see Brathwaite 1993: 282–3), was Louise Bennett. One of the region's most influential writers—particularly as regards her use of Patois—Bennett has been described as the 'only poet who has really hit the truth about her society

through its own language' (Espinet 1990:346). She is famous for her Anancy poems, many of which can be found in *Jamaican labrish*, which means Jamaican Gossip. Anancy is a legendary trickster (spider) in Jamaican folklore, and references to Anancy tend to be more authentic and culturally flavored in WCC than in (Standard) English. An example of an Anancy poem in which the poet switches between WCC and English is Shara MacCullum's poem 'Jack Mandoora: Me No Choose None'.

After Bennett paved the way for WCC usage, many Caribbean writers began to 'explore ways of working the rich ore of dialect in literary contexts' (Burnett 1986:xxv). As a result of the negative associations of the term 'pidgin', the term 'nation language', coined by Barbadian poet, Kamau Brathwaite—another luminary Caribbean poet—is frequently employed in the region's literary circle. This term is believed to aptly echo the functional dimensions served by WCC (Burnett 1986:xxv). Brathwaite (1993) defines it as 'Creole English... the kind of English spoken by the people who were brought to the Caribbean' (260) and 'not the official English' which 'is an imperial language' (259).

Louise Bennett, and other Caribbean writers who employ WCC in their works have helped both promote and sustain this language. In Burnett's words, 'The hybrid tongues which result [in the literature] have an enormous range of nuance and vigor of expression, with the limitation that only the locals catch every resonance' (xxv–xxvi). Noted artists include Orlando Patterson, Michelle Cliff, Lorna Goodison, Merle Collins, Erna Brodber, Patricia Powell, Honor Ford-Smith, Olive Senior, and Pauline Crawford (all Jamaican), and Zee Edgell, Amy Nicholas, Rutheen Taylor, Ushanda Io Elima, and Ziola M. Ellis from Belize. Edgell's *Beka lamb*, co-recipient of the Fawcett Society Book prize in 1992, contains some WCC and was the first Belizean novel to win international acclaim.

Any discussion of WCC would be incomplete without reference to the historical context that engendered WCC, and its popular acceptance. Caribbean Creole is believed to define and authenticate the Caribbean experience (Lamming 1960; Glissant 1989; Gikanda 1992; Benítez-Rojo 1992; Walcott 1993; Hamner 1997). It is no wonder the term *Creole* has many meanings in this region; it is used to describe the language, the people, and the culture (see Brathwaite 1977, 1993; Edgell 1982). To the Caribbean writer, Creole and creolization have acquired vital sociopolitical associations. Creole is conceptualized as 'a new mode of expression and representation' (Gikanda 1992:2) and creolization as a 'counter-discourse away from outmoded and conventional modes of representation associated with colonial domination' (5). In the collection *Caribbean creolization*, some of the region's foremost artists reflect on creolization and their reasons for employing Creole. The discussion opens with a quote from Edouard Glissant:

> Creolization's most manifest symbol is the Creole language. Its genius rests on its being always open.... Thus, creolization carries in itself the adventure of multilingualism along with the extraordinary explosion of cultures. But this explosion does not mean their scattering nor their mutual dilution. It is the violent manifestation of their assented, free sharing (1).

Part One is titled 'Creolization and the creative imagination', and Part Two, 'Creolization, literature, and the politics of language'. The consensus among the contributors is that Creole (including WCC) is a malleable, poetic-license-engendering and liberating indigenous code, free from colonial associations, and best suited to the themes that Caribbeans can relate to. Creolization is defined as 'a cultural process rooted in long-lasting psychological, spiritual, anthropological, and linguistic experiences' (9). Various historical and metaphorical terms are used to describe WCC and its formation, including: *blendings, cradle-hammock, marronage, rhizomatic transcultural exchanges, hybridity, métissage, mediation (contexts), linguistic crossbreeding* borne of plantations, and 'a language of mixed sounds, tonalities, rhythms, and flavors' (10). Some of these descriptors are constructs associated with prominent Caribbean writers. The metaphor of the shipwreck is utilized by some, including Nobel laureate Derek Walcott (1993:11), to describe creolization.

A prime example of an instructional text aimed at eliciting respect for WCC or 'Patois English' (Bennett 1981:i) is *Jamaica maddah goose*. This bilingual collection of nursery rhymes has a clear message, stated in Bennett's introduction:

> Standard English is the official language of the Jamaican businessman, educator, professional and mass media. But it is the mellifluous patois that is the language of the people.... [T]he pictures and, basically, the words are authored by Jamaican art students. Hopefully they reflect what all Jamaican children behold around them. Our only regret is that you can't hear these nursery rhymes spoken. Because patois is a language of melody and rhythm, a language to be heard. It is a vital language, an everchanging language, a language of the street and the scholar. And a language to be appreciated in its own right (i).

Exemplary nursery rhymes in WCC include 'Lickle Miss Julie' (Little Miss Muffet), 'Ole Maddah Hubad' (Old Mother Hubbard), and 'Mary Had Wan Lickle

Lam'. Bennett's mission—to destigmatize WCC—is clear in her poem 'Bans O' Killing', which begins:

> So yuh a de man, me hear bout!
> Ah yuh dem say dah-teck
> Whole heap 'English oat' sey dat
> Yuh gwine kill dialect!

[Translation: So you're the man I've heard about. Ah! You're the one they say uses piles of words and (that) vows to kill Creole!]

The poem continues with the following message, portending a bright future for WCC:

> Dah language wey yuh proud o'
> Wha yuh honour and respeck
> Po' Mass Charlie: Yuh noh know sey
> Dat it spring from dialect!

[Translation: The language that you are proud of, that you honor and respect, Poor Mr. Charlie; you don't (even) know that it came from the dialect you despise!]

References

Alleyne, Mervyn. 1979. Acculturation and the cultural matrix of Creolization. Pidginization and Creolization of languages, ed. by Dell Hymes, 169–86. Cambridge: Cambridge University Press.

Bickerton, D. 1981. Roots of language. Ann Arbor: Karoma.

Collins, Merle. 1998. Angel, London: Seal Pr Feminist Publishers.

Collymore, F.A. 1965. Glossary of Barbadian dialect, 2nd edition. Bridgetown, Barbados: Advocate Co.

Edgell, Zee. 1982. Beka lamb. London: Heineman.

Glissant, E. 1989. Caribbean discourse: selected essays, translated by J. Michael Dash, Charlottsville: University of Virginia Press.

Harney, Stefano. 1996. Nationalism and identity: culture and the imagination in a Caribbean diaspora. London and New Jersey: Zed Books.

Hart, Richard. 1998. From occupation to independence: a short history of the peoples of the English-speaking Caribbean region. Barbados and London: Canoe Press and Pluto Press.

Holm, John, A. 2000. An introduction to Pidgins and Creoles. Cambridge: Cambridge University Press.

James, Louise (ed.) 1999. Caribbean literature in English. London and New York: Longman.

Journal of Pidgin and Creole languages. 1986 to date. Amsterdam and Philadelphia: John Benjamins.

Mais, Roger. 1974. Brother man. London: Heinemann.

McKay, Claude. 1933. Banana bottom, New York: Harper and Brothers.

McWhorter, John, H. 1997. Towards a new model of Creole genesis. studies in ethnolinguistics, Vol. 3. Berlin, New York and Paris: Peter Lang.

Patteson, Richard, F. (ed.) 1998. A critical perspective on new fiction from the West Indies. Boulder and London: Three Continents Press.

Powell, Patricia. 1993. Me dying trial. Ibadan and Gaberone: Heinemann.

Senior, Olive. 1989. Arrival of the snake-woman and other short stories. Harlow, Essex: Longman.

Walcott, Derek. 1993. The Antilles: fragments of epic memory. New York: Farrar, Strauss, and Giroux.

ANITA PANDEY

See also **Arawak; Discourse Analysis; Hawaiian Creole English; Nigeria; Pidgins and Creoles; Krio; Saramaccan; Sranan; Tok Pisin**

Whorf, Benjamin Lee

Along with his mentor, Edward Sapir (1884–1939), Benjamin Lee Whorf is best known today for the principle of linguistic relativity, also called the Sapir–Whorf hypothesis (see 'Sapir–Whorf hypothesis'). Basically self-educated in linguistics and without a Ph.D. (although he did study for a time with Edward Sapir at Yale University starting in 1931), Whorf published extensively on American Indian languages, especially (Uto-Aztecan) Hopi, and also wrote on many other related linguistic topics for a variety of publications. As a result of his research into the structure of these languages, he came to assert that language and culture were intimately related such that one's native language, i.e. its structure or grammar and vocabulary, influences the way a speech community perceives and conceives of its reality.

Whorf's fascination with ciphers, puzzles, and codes as a child as well as his interest in religion led to his reading a French mystic, theosophist, and quasilinguist, Antoine Fabre d'Olivet (1768–1825), whose *The Hebrew tongue restored*, originally published in French in 1815–1816 (translated by Nayan Redfield, Whorf's neighbor, New York: G.P. Putnam's Sons, 1923), exerted a profound influence on the Methodist Episcopal/Theosophist Whorf, about whom he wrote: '[he] cut loose from dead traditional grammar, anticipated the modern concept of the phoneme…, saw the importance of language for logic and what today would be called psychology' (Rollins 1980: 49). Reading another theosophist–linguist helped develop Whorf's ideas even further—Frederick Max Müller (1823–1900), who maintained in his *The science of thought* (1887:295): 'Language is not outside the mind, but is the outside of the mind. Language is very thought as thought is very language.'

Part of the groundwork for the Sapir–Whorf hypothesis was laid by Whorf's work as a fire insurance investigator. During his career, he had the opportunity to analyze many reports as to why fires broke out. He found that workers would use extreme caution when around 'full' drums of gasoline. Just as one might expect, workers were careful not to smoke around 'full' drums. Yet, these same workers when around 'empty' drums of gasoline would sometimes toss lit cigarettes nearby. This caused a violent explosion because an empty drum still contained volatile gasoline vapor. Thus, an 'empty' drum was really much more of a threat than a 'full' one. Using these data, Whorf concluded that the meanings of certain words had an affect on a person's behavior.

It was the painstaking research of both Sapir and Whorf into the grammatical systems of American Indian languages, however, that proved to have the greatest impact on Whorf's subsequent thinking. By predicating their insights into the interrelationships of language and culture on what they had learned from the structures of these languages, the basic idea of language shaping the perceptions of its speakers and providing for them a vehicle so that their experiences and emotions could be placed into significant cognitive categories was given its scientific underpinnings. Generally, Sapir is credited as giving the problem of establishing the link between language and culture its initial formulation (continuing in the tradition of Johann Gottfried Herder, 1744–1803, and Wilhelm von Humboldt, 1762–1835), while Whorf is honored as the one who took this idea and developed it further to include grammar in addition to lexis, thereby making it into a bona fide hypothesis. However, it must be noted that there are really two different versions of the hypothesis. This is understandable when one considers that Whorf did all of his professional writing in the rather short period from 1925 until his untimely death in 1941, and his ideas, quite naturally, were continuously evolving. The strong version of the hypothesis, which is called linguistic determinism, holds that language determines thinking. This position is most difficult to defend primarily because translation between one language and another is possible, and 'thinking' can take place without language at all; e.g. an artist or sculptor can and often does think with his fingers.

Mirroring Sapir's thoughts, Whorf notes in his (1940) 'Science and linguistics':

> We dissect nature along lines laid down by our native languages… We cut nature up, organize it into concepts, and ascribe significances as we do, largely because we are parties to an agreement to organize it in this way—an agreement that holds throughout our speech community and is codified in the patterns of our language. (Carroll 1956:213)

The milder version of the hypothesis is labeled linguistic relativity. This claims that the native language influences thoughts and perceptions. In fact, it was Whorf who coined the phrase 'linguistic relativity', since he always tried to qualify his assertions. In 'Linguistics as an exact science', Whorf maintains:

> …what I have called the 'linguistic relativity principle' … means, in informal terms, that users of markedly different grammars are pointed by their grammars toward different types of observations and different evaluations of externally similar acts of observation, and hence are not equivalent as observers but must arrive at somewhat different views of the world. (Carroll 1956:221)

One of Whorf's most famous essays, 'A linguistic consideration of thinking in primitive communities', asserts that the problem of 'thinking' by so-called 'primitive' peoples is 'approachable through linguistics' (Carroll 1956:65). Further, as linguists have come to fully appreciate only fairly recently, Whorf believed that 'linguistics is essentially the quest of MEANING' (1956:73). Examples are given of things which are relatively easy to say in Hopi but awkward or clumsy to say in SAE (Standard Average European) languages. Whorf concludes that Hopi reveals a 'higher plane of thinking' (ibid.).

In 'Language, mind, and reality', published in India a year after his death, Whorf explains the 'obviative' in Algonkian languages. It is in essence two third persons (Carroll 1956:265), one of which we traditionally refer to as the fourth person. This aids in compact description of complicated situations, for which we should have to resort to cumbersome phraseology. Let us symbolize their third and fourth persons by attaching the numerals 3 and 4 to our written words. The Algonkians might tell the story of William Tell like this: 'William Tell called his[3] son and told him[4] to bring him[3] his[3] bow and arrow…' He remarked that 'such a device would greatly help in specifying our complex legal situations, getting rid of the 'party of the first part'…, 'the party of the second part', the 'aforementioned' or the 'aforesaid' (ibid.).

For those wishing to go to the fountainhead of the debate over language's influence on thought, Whorf's writings still provide for a satisfying journey.

Biography

Benjamin Lee Whorf was born in Winthrop, Massachusetts on April 24, 1897. He graduated from Winthrop High School in 1914, and did his B.S. (1919) in chemical engineering (MIT). In 1920, he began his career as a trainee in fire prevention engineering for the Hartford Fire Insurance Company. He was appointed as a Special Agent for the company in 1928. Eventually,

he became a recognized expert in industrial fire prevention, writing 'Blazing Icicles', which offered a linguistic interpretation of the cause of some fires. He was promoted as Assistant Secretary of the company in 1940.

During the 1930s he regularly visited a Hopi informant in New York City, and five years later had prepared a grammar and dictionary (Unpublished Papers, Yale University). He published much of his research in the major recognized outlets, such as *Language, International Journal of American Linguistics*, and *American Anthropologist*. Although he was offered academic positions, he decided to remain with the Hartford Fire Insurance Company, since he believed it afforded him greater opportunity to do what he wanted. He died of cancer on July 26, 1941.

References

Carroll, John B. (ed.) 1956. Language, thought, and reality: selected writings of Benjamin Lee Whorf. Cambridge, MA: MIT Press.

Gumperz, John J., and Stephen C. Levinson (eds.) 1997. Rethinking linguistic relativity. Cambridge: Cambridge University Press.

Hoijer, Harry (ed.) 1954. Language in culture. Chicago, IL: University of Chicago Press.

Lee, Penny. 1996. The Whorf theory complex. Amsterdam and Philadelphia: John Benjamins.

Rollins, Peter C. 1980. Benjamin Lee Whorf: lost generation, theories of mind, language, and religion. Ann Arbor, MI: University Microfilms International.

Whorf, Benjamin Lee. Unpublished papers. Sterling Memorial Library, Yale University.

ALAN S. KAYE

See also **Sapir, Edward; Sapir–Whorf Hypothesis**

Wittgenstein, Ludwig

Ludwig Josef Johann Wittgenstein was a charismatic Austrian philosopher whose philosophical career developed mainly in England. Wittgenstein's life was a search for a special kind of spirituality by means of philosophical activity. This spirituality is evidenced by some idiosyncratic ethical and even metaphysical ideas in his *Notebooks* 1914–1916 and *Secret diaries*. Wittgenstein's work was very influential in twentieth-century philosophy. He also shaped linguistics —especially pragmatics— mainly through his *Philosophical investigations*, published in 1953. His theory of language-games and his idea that 'the meaning of a word lies in its use' became the basis of certain theoretical frameworks in linguistics. His philosophy of language and his scientific epistemology are also reflected in most of the current social sciences.

The end of the nineteenth and the beginning of the twentieth century saw the arrival of the philosophy of language. Analytical philosophers like Gottlob Frege (1848–1925) conceded a central role to language in philosophical theory, and they begun the construction of a general theory of meaning, which greatly influenced Wittgenstein's thinking. Wittgenstein made two prima facie contradictory contributions to this linguistically oriented philosophy: his *Tractatus logico-philosophicus* (1922) represents an exercise in analytical philosophy (which regards everyday language as an impure vehicle of thought), while his *Philosophical*

investigations (1954) influenced the so-called ordinary language philosophy (which relinquishes the search for an ideal logical language). This is the reason why one traditionally speaks of the 'first' and the 'second' Wittgenstein. However, these 'two' Wittgensteins are reconcilable, as shown below.

Wittgenstein wrote the *Tractatus logico-philosophicus* during World War I, while in a prison camp, and sent it to his teacher and colleague Bertrand Russell (1872–1970). The *Tractatus* was published in England in 1922 with the help of Russell, who also wrote the Prologue. The *Tractatus* is a collection of numbered statements elaborating on some philosophical problems discovered by Frege and Russell. The *Tractatus* expounds a philosophy of language and mathematics and is a reflection on the nature of life and philosophy. For Wittgenstein, philosophy is mainly an activity, not a doctrine, and so his philosophy must be understood through his works.

Wittgenstein's philosophy, as presented in the *Tractatus*, relates three elements: reality, thought, and language. Reality exists, thought is the interpretation of reality, and language is the expression of thought. Wittgenstein calls the logical identity between the structure of reality, the structure of thought, and the structure of language 'isomorphism'. For him, logic is the ideal language, but only a tool for description. Language is used to describe facts, although there

remain things that cannot be said. Therefore, the *Tractatus* reaches the now famous conclusion: 'what we cannot speak about we must pass over in silence'. Thus, Wittgenstein studies the limits of language and correspondingly the limits of oneself: 'The limits of my language are the limits of my world' (*Tractatus* 5.6).

The *Tractatus* inspired some meetings of a group of philosophers referred to as the 'Vienna Circle', who developed the so-called *logical positivism* or *neopositivism*. The principles of particular appeal to the Vienna Circle were the analytical characteristics of logic, the claim that sentences need to be verifiable to be true and the idea that philosophy clarifies truth instead of discovering it.

The blue and brown books mark a transition from the *Tractatus* to the *Philosophical investigations*, that is, from the *first* to the *second* Wittgenstein. The *Philosophical investigations* were published posthumously, and are a collection of Wittgenstein's work between 1941 and 1949. The concept of philosophy, the rules of language, and the theory of language games are its main topics. It is precisely the concept of 'language game' that reconciles the *Tractatus* and the *Philosophical investigations* as discussed below.

The *Philosophical investigations* shares with the *Tractatus* that philosophy is seen as an activity, and this activity consists of the analysis of language. However, in the *Tractatus*, Wittgenstein was only interested in the logical truth or falsity of sentences. In the *Investigations*, on the other hand, he was willing to look for the meaning of words as expressed in their everyday use. Wittgenstein thus acknowledged that we are primarily users of language and only secondarily analysts of language.

For Wittgenstein, the meaning of words and propositions is identifiable in their use in language, and this use is shaped in 'language games'. The concept of language games highlights the idea that language use is rule-governed, i.e. the game metaphor is a model of the rule-based aspects of language. Wittgenstein uses the term 'game' broadly:

'23. […] Review the multiplicity of language-games in the following examples, and in others:
Giving orders, and obeying them-
Describing the appearance of an object, or giving its measurements-
Constructing an object from a description (a drawing)-
Reporting an event-
Speculating about an event-
Forming and testing a hypothesis-
Presenting the results of an experiment in tables and diagrams-
Making up a story; and reading it-
Play-acting-
Singing catches-
Guessing riddles-
Making a joke; telling it-
Solving a problem in practical arithmetic-
Translating from one language into another-
Asking, thanking, cursing, greeting, praying.'

Because the game metaphor thus extends to all rule-governed uses of language, formal language or the language of logic can also be viewed as a language game. Hence, the *Tractatus* is also reflected on in the *Philosophical investigations*. Wittgenstein's work is thus a language game itself, which highlights the self-reflexive perspective a Wittgensteinian approach has to take.

Understanding the meaning of a word requires a knowledge of linguistic and extralinguistic facts, and Wittgenstein asserted that the analysis of language use is as important as logical analysis. This represented a 'linguistic turn' in the philosophy of language, and the resulting requirement of critically reflecting on terminology now forms an important part of the scientific epistemology of many social sciences. Since the *Philosophical investigations* with its 'language games' introduced a rule-based view of discourse, it can be viewed as the philosophical basis of pragmatics. Members of the Oxford School such as G. Ryle, P.F. Strawson, and notably J.L. Austin, who developed the theory of 'speech acts', were heavily influenced by Wittgenstein's ideas. Thus, after philosophy turned to language, linguistics turned to philosophy: 'this was the "pragmatic turn", which led to a proliferation of communicative and social studies of speech in action and action in speech' (Nerlich and Clarke 1996:6).

Biography

Born in Vienna, Austria, on April 26, 1889 to a wealthy family, Ludwig Wittgenstein was educated in an artistic and intellectual environment at home until he was 14. From 1903 to 1906 he studied mathematics and natural sciences in Vienna, and from 1906 to 1911 he studied Mechanical Engineering in Berlin and Manchester. He went to Cambridge in 1911 to attend lectures given by Bertrand Russell on mathematical logic, but left Cambridge somewhat disappointed with the knowledge he gained there and went to Skjolden, Norway. He joined the Austrian army from 1914; in 1918 he became prisoner of the Italians. During these war years he wrote the *Tractatus*, which he sent to Russell while he was in a prison camp. Released in 1919, he became primary teacher in 1920 and taught until 1925. After fulfilling various posts, he returned to academic life in 1929: he taught at Trinity College. That year he submitted the *Tractatus* as his doctoral thesis. In 1939 he became Professor of Philosophy, and remained at Cambridge except for the World War

II period. In 1947 he retired to Ireland, and in 1949 came back to Oxford because of his illness. He continued working until he died in Cambridge on April 29, 1951. A large part of his works was published posthumously.

References

Anscombe, G. Elizabeth. 1971. An introduction to Wittgenstein's Tractatus. Philadelphia: University of Pennsylvania Press.

Baum, Wilhem. 1985. Ludwig Wittgenstein. Berlin: Colloquiun Verlag GmbH.

Kripke, Saul A. 1982. Wittgenstein on rules and private language. Cambridge, MA: Harvard University Press.

Malcolm, Norman. 1958. Ludwig Wittgenstein. A memoir. Oxford: Oxford University Press.

Nerlich, Brigitte, and David Clarke. 1996. Language, action and context. The early history of pragmatics in Europe and America, 1780–1930. Amsterdam and Philadelphia: John Benjamins.

Specht, E.K. 1963. The foundations of Wittgenstein's late philosophy. Manchester: Manchester University Press.

Winch, Peter (ed.) 1969. Studies in the philosophy of Wittgenstein. London: Routledge and Kegan Paul.

Wittgenstein, Ludwig. 1961. Notebooks 1914–1916. Oxford: Blackwell.

———. 1922 and 1961. Tractatus logico-philosophicus. London: Routledge and Kegan Paul, reprinted 1993.

———. 1958. The blue and brown books. Oxford: Blackwell [Harper Collins Publishers, 1976].

———. 1953. Philosophical investigations. Oxford: Blackwell, reprinted 1996.

———. 1973. Letters to C.K. Ogden. Oxford: Blackwell.

———. 1986. Philosophical grammar. LA: University of California Press.

ELENA BATTANER-MORO

See also **Pragmatics**

Wolof and Atlantic Languages

Wolof is an Atlantic language of the Niger-Congo phylum, spoken primarily in Senegal and the Gambia on the Atlantic coast of West Africa. Wolof serves as a lingua franca in Senegal, where approximately 40% of the population speak it as a first language and 45% speak it as a second language. Extrapolating from the 1988 census figures, there are currently at least three million native speakers of the language, and at least as many who speak it as a second or third language, making for a total of no less than six million speakers.

The Atlantic languages are spoken almost exclusively along the West African coast from the Senegal river that forms the border between Mauritania and Senegal southward to Sierra Leone and Liberia, generally abutting languages of the Mande family toward the east. Atlantic languages with the greatest numbers of speakers include Fula, which is spoken across the Sahel; Wolof and Seereer, both spoken primarily in Senegal; and Temne, spoken in Sierra Leone. Fula (also *Peul* in French from Wolof *Pël*, and Fulani in various languages, including English, from Hausa *Fùláánìì*) is known by its speakers as Pulaar (or Pular) in the western dialects and Fulfulde in the eastern ones. It is spoken by several million people across a noncontiguous stretch of West Africa. Fula speakers are found from Mauritania and Senegal in the west to Chad and Sudan in the east, including significant populations in Guinea, Mali, Burkina Faso, Nigeria, Niger, and Cameroon. The different dialects of Fula are for the most part mutually intelligible, although influence from neighboring languages has resulted in some significant lexical differences between dialects. Contemporary research confirms Koelle's (1854) hypothesis that places the Fula-speaking heartland in the Fouta Toro region of northern Senegal, with subsequent eastward expansions of Fula-speaking peoples, many of whom were and still are cattle-herding pastoralists. Seereer is spoken by approximately seven hundred thousand people concentrated in the western and central areas of Senegal, centered on the historical region of Siin (Sine), and in parts of the Gambia. Most Seereer speakers speak Wolof as a second language. Temne is one of the major languages of Sierra Leone, spoken in the central part of the country, mostly to the north of Freetown, and current estimates, although possibly low, are that it has approximately one million two hundred thousand speakers.

Atlantic was recognized as a group of related languages as early as Koelle's (1854) *Polyglotta Africana*, where it goes by the name of North-West Atlantic. Westermann (1927) renamed the language group West Atlantic, a name retained by Greenberg in his classification of African languages published in 1963. The name was subsequently shortened to Atlantic by Doneux (1975). Most contemporary classifications of Atlantic consider it to be an early branching of Niger-Congo, concurrent with or subsequent to the Mande branching. Until Greenberg's classification, and largely

as a result of using cultural and ethnic criteria rather than linguistic ones, Fula had been classified apart from the Atlantic languages. Linguistic evidence, however, showed that Fula was clearly an Atlantic language with a strong resemblance to Seereer. Most contemporary classifications of Atlantic, including Wilson (1989), are based on or respond to Sapir's (1971) lexicostatistical study of the languages. Comparing 100 core lexical items across languages, which revealed rather low correspondences, Sapir nevertheless posits a Northern Branch, Southern Branch, and a single isolated language, Bijago (or Bijogo), spoken on the Bijago Islands off the coast of Guinea Bissau. Sapir's Northern Branch of Atlantic is comprised of five subgroups: (A) Senegal languages (Fula and Seereer, Wolof); (B) Cangin languages (Non, Ndut, Lehar, Safen, also sometimes called Seereer-Non, Seereer-Ndut, etc.); (C) Bak languages (Diola cluster, Manjaku and Papel, Balanta); (D) Eastern Senegal — Portuguese Guinea (now Guinea Bissau) languages (Tenda group: Tanda, Basari, Bedik, and Konyagi; Biafada and Pajade; Kobiana and Banhum; and (E) a nameless group (Nala, Mbulungish, and Baga Mboteni). The Southern Branch is comprised of three subgroups: (A) Sua; (B) Mel languages (Temne and Baga Koba, Sherbro, Mmani, Krim and Kisi, Gola); and (C) Limba. Recent unpublished work on the isolate, Bijogo, suggests that it may be the most conservative Atlantic language, especially since its noun class markers most closely resemble those of the Bantu languages. The low scores found by Sapir for cognates between Atlantic languages are corroborated by Bennett and Sterk (1977), who find that the Atlantic languages generally show no more genetic affinity with each other than they do with the rest of the Niger-Congo phylum. Although grammatical and lexical similarities are to be found within various subgroups, to date linguists have not found any evidence of shared innovation common to the Atlantic languages, leaving their coherence as a genetically related family a matter of convention rather than fact.

Salient characteristics of the Atlantic languages include an SVO word order, extensive noun class systems, and verbal extensions, as well as a perfective/imperfective aspectual distinction in the verbal system, a syntactically encoded focus system, and verb serialization. These features are not unique to Atlantic languages but are, in fact, common to the Niger-Congo phylum as a whole. The trait most closely associated with, although not unique to, the Atlantic languages is morphologically conditioned stem-initial consonant mutation, which is found on nouns, adjectives (for those languages that have them), and verbs throughout the language group.

The consonantal inventories of Atlantic languages generally distinguish four places of articulation: labial, alveolar, palatal, and velar. Some languages, including Wolof and Seereer, also have one or more uvular consonants. Prenasalized stops and implosive consonants are common, although Wolof does not have the latter, and word-initial voiceless prenasalized stops have all but disappeared from the language, being retained only in place names (such as *Mpal*) or by older rural Wolof speakers. Seereer is phonetically quite unusual in having a series of both voiced and voiceless implosive stops in three places of articulation: bilabial, alveolar, and palatal. A basic five-vowel system is augmented with [+/−ATR] (advanced tongue root) contrasts for mid-vowels in many Atlantic languages. These are phonemic in Wolof, as evidenced by pairs such as *reer* [−ATR] 'dinner' and *réer* [+ATR] 'to be lost'. Wolof also makes an [ATR] contrast between the low vowel [a], which is [−ATR], and a central schwa-like vowel written as ë. Many Atlantic languages, including Wolof, contrast both consonant and vowel length; Seereer contrasts only the latter. While almost all Niger-Congo languages, including most Atlantic languages, are tonal, the northern Senegal languages, Fula, Seereer, and Wolof, are not, although a high-pitch accent is associated with ideophones.

Morphologically conditioned consonant mutation is found stem initially in verbs and nouns in a great number of Atlantic languages. In noun stems, initial consonants may have up to three homorganic variants conditioned by noun class. The variants are known as grades, following Arnott (1970). In Fula, these take the form of a continuant, a stop, and a prenasalized stop as in the singular (Class 1), plural (Class 2), and diminutive plural (Class 21) words for 'man' (*gorko, worɓe, ⁿgoron*) and 'woman' (*debbo, rewɓe, ⁿdewon*). Seereer exhibits a similar type of consonant mutation, as well as one that shows alternations between a voiced stop, voiceless stop, and prenasalized stop. The word for 'man' in its singular (Class 1), plural (Class 2) and diminutive singular (Class 12) forms (*okoor, goor, oⁿgoor*) illustrates the latter type of mutation, while the same forms for the word for 'woman' (*otew, rew, oⁿdew*) illustrate the former. This type of consonant mutation in the Atlantic languages has been analyzed as the historical result of the erosion of noun class prefixes which had a phonological effect on the stem-initial consonant. Remnants of prefixes are seen in some Atlantic languages like Seereer, providing incontrovertible evidence that Atlantic languages, like other Niger-Congo languages, and particularly Bantu languages, at one time had a complete set of noun class prefixes. Such prefixes eroded phonologically over time only to be renewed in many cases, such as Fula, by suffixes. Wolof exhibits an advanced stage in the erosion of noun class morphology: for the most part, nouns are not marked for class so that their singular

and plural forms are identical. A few lexicalized exceptions to this generalization, including the words for 'eye/s' (*bët/gët*) and 'person/people' (*waa/gaa*), provide evidence of a once robust inflectional system. Consonant mutation is nevertheless quite productive in other areas of the grammar, including diminutive formation, which requires a prenasalized form, and verb to noun derivation.

Wolof has ten noun classes: eight singular and two plural. Of the singular noun classes, one, the s-class, serves a second function as a diminutive class. A noun class marker appears in the form of a single consonant on nominal dependents such as determiners and relative particles. The word for 'person' (*nit*) is in the k-class, and the word for 'bird' (*picc*) is in the m-class, hence the following forms: *nit ki* 'the person'/ *picc mi* 'the bird': *nit kële* 'that person' / *picc mële* 'that bird'; *nit ku baax* 'person who is good'/ *picc mu baax* 'bird that is good'. A problematic aspect of Wolof class assignment, remarked upon by several scholars, is the tendency for the class marker to repeat the initial consonant of certain nouns, in a process akin to consonant harmony or reduplication. Phonologically conditioned agreement poses a substantial problem for linguistic theory in general and the lexicalist hypothesis in particular, since syntax should not have access to the phonology. A second noted trend in the evolution of the Wolof noun class system is a tendency toward assigning nouns to the default b-class. The origins of this tendency are to be found in a combination of morphological and sociolinguistic factors (Irvine 1978; Mc Laughlin 1997).

African languages in general, including Atlantic languages, have had a profound effect on contemporary linguistic theory. Consonant mutation in the Atlantic languages, and particularly in Fula, has substantially informed underspecification theory and, more recently, theories of featural affixation. The interaction of consonant mutation and reduplication in Seereer-Siin has had implications for correspondence theory; and the facts of ATR vowel harmony in Wolof, which has vowels that are both opaque and neutral to the process, have also been instrumental in theoretical advances in that area. The clause structure of Wolof has also provided evidence for the projection of a sigma phrase within the principles and parameters framework.

The sociolinguistics of Wolof are of particular interest. Wolofization, or the spread of Wolof as a lingua franca in Senegal, has been increasing steadily since the colonial period when it was used widely in the coastal cities. Wolofization is an urban phenomenon, and it is the specifically urban variety of Wolof that continues to spread, even to the point where it has replaced Portuguese creole as the predominant language of Ziguinchor, the largest city in the Casamance region of southern Senegal. Urban Wolof is substantially different from dialects spoken in the predominantly rural Wolof heartland. Its main characteristics are extensive lexical borrowing from French, the former colonial language and current official language of the country, and the reduction of the noun class system toward the default b-class. Urban Wolof has an ambiguous status. On the one hand, as the language of the capital it holds a certain prestige, especially among youth, and its mastery implies a certain urban sophistication. On the other hand, it is recognized as being a hybrid language and looked down upon as 'impure', especially by those who are not ethnically Wolof.

In the domain of discourse, elaborate speech in Wolof is associated with griots or verbal artists who are of low social status. Irvine (1978) has documented the avoidance of elaborate speech as an attempt to manipulate social status.

Finally, although there is an official Wolof orthography in the Roman script, it is not widely used. The preferred writing system is an *ajami* version known as *wolofal*, which is a modified Arabic script, made popular through widespread informal Islamic education and attendance at Qur'anic schools.

References

Arnott, D.W. 1970. The nominal and verbal systems of Fula. Oxford: Clarendon Press.

Bennett, Patrick R., and Jan P. Sterk. 1977. South Central Niger-Congo: a reclassification. Studies in African Linguistics 8. 241–73.

Childs, G. Tucker. 1995. A Grammar of Kisi: a southern Atlantic language. Berlin and New York: Mouton de Gruyter.

Doneux, Jean L. 1975. Hypothèses pour la comparative des langues atlantiques. Africana Linguistica 6. 41–129.

Greenberg, Joseph H. 1963. The languages of Africa. Bloomington: Indiana University Center in Anthropology, Folklore and Linguistics, and The Hague: Mouton.

Irvine, Judith T. 1978. Wolof noun classification: the social setting of divergent change. Language in Society 7. 37–64.

Ka, Omar. 1994. Wolof phonology and morphology. Lanham, MD: University Press of America.

Koelle, Sigismund W. 1854. Polyglotta Africana. London: Church Missionary House.

Mc Laughlin, Fiona. 1997. Noun classification in Wolof: when affixes are not renewed. Studies in African Linguistics 26(1). 1–28.

Sapir, J. David. 1971. West Atlantic: an inventory of the languages, their noun class systems and consonant alternation. Current trends in linguistics: linguistics in sub-Saharan Africa, ed. by Thomas Sebeok. The Hague: Mouton.

Westermann, Diedrich. 1927. Die westlichen Sudansprachen und ihre Beziehungen zum Bantu [The western Sudan languages and their relationship to Bantu]. Berlin: Walter de Gruyter.

Wilson, W.A.A. 1989. Atlantic. The Niger-Congo languages, ed. by John Bendor-Samuel. Lanham, MD: University Press of America.

FIONA MC LAUGHLIN

See also **Niger-Congo**

Word

Although many language users intuitively know what a 'word' is, an unequivocal and clear-cut linguistic definition cannot be given. The concept of 'word' is highly ambiguous and refers to different phenomena. On the one hand, words are language units at the intersection of two linguistic levels of description: morphology (which is the arrangement of meaningful sound sequences, or morphemes, in words) and syntax (which is the arrangement of words in sentences). Accordingly, lexicology is the linguistic branch that deals with words, which thus implies both morphological and syntactic aspects. On the other hand, words are regarded as basic units in the psychological reality of language acquisition, production, and processing: words in the mind. An attempt to produce a system reflecting the versatility of the notion of 'word' would need to describe words as both 'linguistic structures' (although of different kinds) and as parts of the mental lexicon.

The linguistic notion of 'word' captures three fundamentally different, but closely related, concepts that will be indicated in the following as word$_1$, word$_2$, and word$_3$. To begin with, so-called 'word forms' (i.e. word$_1$) refer to the physical substance of words in spoken and written language: phonological word forms in speech, and orthographic word forms in writing. For example, /bɔɪ/ and /bɔɪz/ are two phonological word forms, i.e. specific combinations of sounds in the phonic medium. On the other hand, *boy*, *boy's*, *boys*, and *boys'* represent four corresponding orthographic word forms, or combinations of letters (and apostrophes) in the written medium. The six word forms provide the inventory of possible word-form realizations of one underlying abstract unit, the lexeme *BOY*.

The lexeme (word$_2$), usually given in capitalized letters, is what all the actual word forms have in common at an abstract level. It is this abstract unit of a lexeme that is at the basis of a dictionary since all word forms of one and the same lexeme are subsumed under the same dictionary entry.

While word forms represent merely physical forms in terms of sounds or letters, lexemes—in a first approximation—are examples of linguistic signs. That is to say, a lexeme connects potential forms (i.e. the signifier) to an abstract meaning encoded by these forms (i.e. the signified), such as the phonological word forms /teɪbl/ and /teɪblz/ (i.e. the signifier) to the meaning of 'piece of furniture consisting of a flat top supported on one or more legs' (i.e. the signified). This arbitrary, but fixed, form-meaning relation is inherent to the lexeme as a linguistic sign. However, a lexeme actually forms a set of several linguistic signs: some word forms of a lexeme may be linked to specific (sub-) meanings of the lexeme and vice versa. For example, all orthographic word forms of the verb YIELD, i.e. *yield*, *yields*, *yielded*, *yielding*, are linked to the possible meanings 'give way' and 'lead to'. On the other hand, a third meaning of the verb, 'produce', is strongly associated with the first two aforementioned word forms. To be able to refer to such subsets of word forms with a specific range of meaning, the concept of 'lexical unit' or 'lexical linguistic sign' has been introduced.

Apart from word forms and lexemes (including lexical units), the concept of 'word' can also be approached from a grammatical perspective. A grammatical word (i.e. word$_3$) is immediately relevant to syntax and has specific morphological features. The distinction between word forms and grammatical words is important because one word form may represent different grammatical words (a phenomenon called 'syncretism'). For example, the phonological word form /bɔɪz/ can be regarded as a realization of the plural form, the singular genitive, and the plural genitive. In applying the concept of grammatical words, linguists are particularly interested in the question of how words, as minimal syntactic units, are arranged in grammatical structures. Irrespective of the specific grammatical principles at work, grammatical words have a certain ability to change position in a sentence (i.e. 'positional mobility'). Thus, it is often possible to shift a grammatical word, in its entirety, to a different position in a given sentence without jeopardizing the grammatical integrity of that sentence. It goes without saying that in English, which has a comparatively fixed word order, the overall degree of positional mobility is lower than, for example, in German, which allows for much more word-order variation. In terms of their internal structure, however, grammatical words are stable, in that the sequence of morphemes is, in principle, not subject to variation: for instance, *boy-ish* is a possible adjective in English, but the reversed sequence, i.e. **ish-boy*, is not permitted.

Words also play an important role in the mental processes that underlie language use. It is reasonable to assume that words are stored in the mind in a highly structured way. This orderliness of the complex mental lexicon permits the amazing speed of lexical retrieval in natural speech. It is widely accepted that the mental

lexicon presumably does not list each and every word in a random order. Rather, a list of lexical structures (words, word groups, word components) is supposed to be complemented with a set of rules and principles on how to combine these structures. This concept avoids, for instance, a listing of different words ending in the same suffix as separate entries: for example, *boyish* and *girlish* are derived from applying the same combinatorial rule to the base entries *boy* and *girl*, respectively. Such systematic combinations presuppose a mental capacity for analyzing complex words in speech processing, which is subsumed under the notion of 'morphological parsing'. In the aforementioned example, morphological parsing would refer to speakers' ability to strip the suffix *-ish* and to identify a general rule that makes it possible to derive an adjective from a noun by adding this suffix to the noun.

The importance of words for the linguistic competence in the mind and the highly structured design of the mental lexicon are also corroborated by research into language acquisition. The surprisingly rapid development of child language is to a large extent based on the acquisition of thousands of words in a comparatively short time. At a more or less specific stage of this development, children tend to overgeneralize morphological rules. For instance, they add the regular past tense morpheme (*-ed* in English) to all regular verbs (*learn* becomes *learned*) and irregular verbs (*go* becomes **goed*). This phenomenon reflects the tendency for a child's mental lexicon to comprise both basic elements (*learn*, *go*, *-ed*) and so-called rules of morphological parsing to combine these elements and analyze the combinations (verb_present + *-ed* = verb_past).

Lexical disorders, too, shed light on the structure of the mental lexicon. Aphasia—a condition of speechlessness that stroke patients often display when specific areas of the brain are affected—is a compelling example of such a disorder. Broca's aphasia, for example, leads to 'telegraphic speech', or the omission of function words (e.g. articles like *the*, prepositions like *of*), whereas patients suffering from Wernicke's aphasia show difficulty in lexical selection (i.e. they very often do not produce the words they want but other words with different meanings, which often renders their speech unintelligible).

In the light of their structural and mental significance, it is reasonable to assume that words are the most basic linguistic units. However, the overall importance of words for language to function is largely at odds with the yet unresolved problem of word identification. What is relatively simple in written language, where words are usually separated by spaces from each other (though not the case in all languages), is far more complicated in spoken language. Several

attempts have been made to establish criteria for a reliable identification of words in the continuous stream of speech sounds. For example, words are said to be demarcated by positions at which pausing is, at least, possible. Another suggestion is to look for those stretches of sounds that could stand meaningfully on their own. Although these two approaches and others may be useful for the majority of words, other cases cannot be clarified, such as compounds, which might be considered as one word or as several words.

In all languages, linguists set out to group words into word classes. Generally, one can distinguish between 'open' and 'closed' word classes. Open word classes can always have new words freely added to them; in most languages, these classes include nouns, verbs, adjectives, and adverbs. In contrast, usually no new members can be added to closed word classes. In English, for example, prepositions (e.g. *under*), conjunctions (e.g. *because*), pronouns (e.g. *mine*), determiners (e.g. *the*), and numerals (e.g. *first*) are the primary closed word classes. The basis for the categorization of words into classes is provided by their morphosyntactic behavior. Nouns, for instance, are almost always able to take a plural form (disregarding exceptions). As with word identification, word class categorization is not straightforward in all cases. An important example in English is, perhaps, the word class of adverbs. This class contains members that are extremely dissimilar in structural terms, such as the prototypical adverb *rapidly* and the negative particle *not*.

Words have an internal structure themselves and are at the same time the basic units of larger structures. The former aspect refers to the domain of 'derivational morphology', i.e. the linguistic description of processes and phenomena of word formation. The latter aspect captures the fact that words play a key role in syntactic structures, which they enter in a principle-guided fashion. Moreover, a given word in a specific position tends to predict, to a large extent, the range of words possible both before and after it. If the probability of two words occurring next to each other is significantly high, this phenomenon is called 'collocation' (from *co-location*).

With collocations, corpus linguistics has definitely broken new ground. In searching very large computer collections of text (so-called corpora) for co-occurrences, it is now possible to base the concept of collocation on an empirical foundation. By applying such corpus linguistic methods, it can be statistically determined to what extent a given word predicts another word in its immediate neighborhood. For example, *strong* and *argument* co-occur in significant frequency in authentic English language use. Their combination is thus not of a random nature but constitutes a collocation, i.e. a lexical co-selection. Furthermore, specific

words tend to occur in a restricted range of grammatical patterns (a phenomenon called 'colligation', from *co-ligation*). For example, *naked eye* almost exclusively occurs in the grammatical pattern 'verb/adjective + preposition + definite article + *naked eye*' as in *visible to the naked eye*. Collocations and colligations reveal that lexical choices and grammatical choices are not independent of each other but that lexis and grammar are intricately intertwined. One could argue that the traditionally established distinction between lexis and grammar should eventually be overcome altogether and be replaced with an integrated lexicogrammar (or a 'pattern grammar').

The linguistic level of words, i.e. lexicology, is of particular interest for dictionary makers or lexicographers. They take an inventory of the words of a language but, by definition, are bound to fail to give an exhaustive account of all words in existence at a given time. Although it is impossible to list all words in a given language, the use of dictionaries has led to a high degree of standardization in orthographic and phonological word forms; they are, of course, a helpful tool for language users in general and language learners in particular. Dictionaries should include orthographic, phonological, grammatical, and semantic information about words, thereby strongly resembling the composition of the mental lexicon as hypothesized by psycholinguists. One ought to bear in mind, however, that the mental lexicon is based on a list of entries and rules to combine them, whereas dictionaries usually contain only an alphabetical list of entries. Furthermore, large dictionaries represent enormous and idealized word lists of which no single native speaker has command: they approximate the so-called 'magnavocabulary' of a language in its entirety, but do not correspond to the specific vocabulary that an individual speaker has at his or her disposal.

The object of inquiry for lexicographers and lexicologists is in constant flux. Productivity and imagination in human languages are perhaps most obvious at the level of words.

References

Aitchison, Jean. 1994. Words in the mind: an introduction to the mental lexicon. Oxford: Blackwell.

Bauer, Laurie. 1983. English word-formation. Cambridge: Cambridge University Press.

Coates, Richard. 1999. Word structure. London: Routledge.

Cruse, D. Alan. 1986. Lexical semantics. Cambridge: Cambridge University Press.

Esser, Jürgen. 2000. Corpus linguistics and the linguistic sign. Corpus linguistics and linguistic theory: Proceedings of the 20th ICAME conference, ed. by Christian Mair and Marianne Hundt. Amsterdam: Rodopi.

Fischer, Roswitha. 1998. Lexical change in present-day English: a corpus-based study of the motivation, institutionalization, and productivity of creative neologisms. Tübingen: Narr.

Hunston, Susan, and Gill Francis. 2000. Pattern grammar: a corpus-driven approach to the lexical grammar of English. Amsterdam: John Benjamins.

Katamba, Francis. 1994. English words. London: Routledge.

Lipka, Leonhard. 1990. An outline of English lexicology: lexical structure, word semantics, and word-formation. Tübingen: Niemeyer.; 2nd edition, 1992.

Marchand, Hans. 1969. The categories and types of present-day English word-formation: a synchronic–diachronic approach. München: Beck.

Stubbs, Michael. 2001. Words and phrases: corpus studies of lexical semantics. Oxford: Blackwell.

Wierzbicka, Anna. 1985. Lexicography and conceptual analysis. Ann Arbor: Karoma.

JOYBRATO MUKHERJEE

See also **Aphasia; Corpus Linguistics**

Word Order

Sentences consist of words that are combined in certain patterned ways to convey meaning. One important dimension of this system of combination is linear order: the order in which words are combined to form phrases and sentences is fundamental to the syntactic structure of any language.

The importance of word order in English, for example, can be seen in the following sentences, in which an asterisk indicates that an order is ungrammatical:

(1) *A dog bit the man.*

(2) *The man bit a dog.*

(3) **Bit a the man dog.*

Only the word order change between (1) and (2) signals a crucial meaning change, and (3) shows that words

cannot be combined haphazardly, but must be ordered by rules. Word order is just as important to the structure of separate phrases as it is to a whole sentence. In English, a long nominal phrase such as *these three new white cotton sheets* is grammatical, but it becomes unintelligible or odd if the ordering of even one word is altered: **new these three white cotton sheets*.

The rules governing word order are specific to each language and may show considerable variation. Nevertheless, research has identified certain universal word order patterns. Typological studies of hundreds of languages show that each one may be classified into one of three basic word order types according to how the three main syntactic constituents—subject (S), verb (V), and object (O)—are ordered within a basic declarative sentence. According to studies by Greenberg (1966) and Tomlin (1986), the most common word order type is SOV, which is found in 37–45% of the languages studied. This order is basic in languages such as Japanese, Turkish, and Chinese. The SVO order of languages such as English, French, and Spanish is nearly as common, occurring in 35–43%. The VSO order found in languages such as Irish, Maori, and Berber is the rarest, with estimates varying from 9% to 20%.

Logically, there are at least three other possible word order types—VOS, OSV, and OVS. However, these orders are virtually nonexistent. Their absence is often interpreted as a universal preference for ordering subjects before objects. One possible functional explanation for this preference may be that it reflects the way in which humans arrange information cognitively and that subjects are often more salient cognitively than objects. The relative rarity of the VSO order suggests that there is a preference for keeping the verb and object together as a unit, as they are in the SVO and SOV orders.

Some researchers, for example Lehmann, instead simplify the three basic word order types into two, making VSO a variation of an underlying SVO structure. In this view, all languages use either a basic VO or OV order. Hawkins and Vennemann have also expressed this as a general head-dependent (i.e. VO) vs. dependent-head (OV) order, which can be seen to govern the ordering of all types of phrases within one language.

Phrase-internal orderings among word categories such as nouns (*house, dog,* etc.), determiners (*the, a,* etc.), adjectives (*green, dark,* etc.), possessives (*my, your,* etc.), and adpositions (*in, for, of,* etc.) show interesting correlations with the three basic sentential word orders. For example, VSO languages tend to use prepositions (*in* in the phrase *in the house*), whereas SOV languages use postpositions (literally translated, *the house in*). Similarly, VSO languages follow noun–adjective order (*house white*), whereas SOV languages follow adjective–noun order (*white house*). Such correlations also support the OV vs. VO typology, although there are languages that contain exceptions.

Languages differ in the degree to which they allow or exploit variations from the basic SVO, SOV, or VSO word order. Variations are typically used to convey a certain style or focus, emphasize a particular word or phrase, or signal that a sentence has a special discourse function, such as a question or command. It is not uncommon to find that one word order is standard in finite (tensed) main clauses and that another is the norm in infinitive (untensed) clauses. Adverbs (such as *luckily*), which are optional elements, are particularly flexible in their possible placements. In terms of language processing, the very first and last positions in each sentence are significantly prominent positions. Often a phrase will be ordered first or last to emphasize and focus attention on it. Very long phrases, such as *the three young men who washed the car for me last Saturday,* are often positioned last or at least moved rightward, possibly because this makes them easier to process. The first and last positions are also the most common ones for special question-marking particles in those languages that use such particles.

The second position in a sentence ('Wackernagel position') also seems to have special significance. One group of languages normally classified as SVO shows a basic word order in which the first position may actually be occupied by any type of phrase (not just the subject), as long as the verb consistently appears in second position. This word order pattern is known as verb-second or V2 order and has been studied extensively in research on the Scandinavian languages, German, Dutch, and Yiddish. Second position also seems to be a preferred position in many languages for the placement of clitics, i.e. elements that cannot stand alone but must attach to an independent 'host' word.

Finally, some languages, such as Polish, may be classified as showing 'free word order' at the sentence level. That is, it is possible to combine phrases in many different ways and still convey essentially the same meaning. Also, some languages allow for discontinuous constituents, in which the words of one phrase may be separated rather than appearing as one continuous string of words. These ordering phenomena present interesting challenges to existing theories of both linear and hierarchical syntactic structure.

In summary, it is most common for languages to identify one basic word order for the subject, object, and verb in each sentence, but to allow for some stylistic variations. In either case, word order is rule governed and shows some universal tendencies that may reflect possible constraints on how humans process language.

References

Baker, Mark C. 2001. The natures of nonconfigurationality. The handbook of contemporary syntactic theory, ed. by Mark Baltin and Chris Collins. Oxford: Blackwell.

Greenberg, J.H. 1996. Some universals of grammar with particular reference to the order of meaningful elements. Universals of language, 2nd edition, ed. by J.H. Greenberg. Cambridge, MA: MIT Press.

Haider, Hubert, and Martin Prinzhorn (eds.) 1986. Verb second phenomena in Germanic languages. Dordrecht: Foris.

Hale, Ken. 1992. Basic word order in two 'Free Word Order' languages. Pragmatics of word order flexibility, ed. by D. Payne. Amsterdam: John Benjamins.

Halpern, Aaron, and Arnold Zwicky (eds.) 1996. Approaching second: second position clitics and related phenomena. Stanford: CSLI Lecture Notes.

Hawkins, John A. 1983. Word order universals. New York: Academic Press.

Lehmann, Winfred. 1973. A structural principle of language and its implications. Language 49. 47–66.

Siewierska, Anna. 1988. Word order rules. London: Croom Helm.

Tomlin, Russell S. 1986. Basic word order: functional principles. London: Croom Helm.

Vennemann, Theo. 1974. Analogy in generative grammar: the origin of word order. Proceedings of the Eleventh International Congress of Linguists, ed. by L. Heilmann. Bologna: Il Mulino.

SHEILA DOOLEY COLLBERG

See also **Greenberg, Joseph Harold; Syntactic Typology; Wackernagel, Jacob**

Word Sense Disambiguation

A word is semantically ambiguous if it has more than one sense. Word sense disambiguation (WSD) is the process of deciding which sense is correct in a given context. The process of WSD is illustrated by Rayson and Wilson's SEMSTAT (Thomas and Wilson 1996), a semantic tagger that reads in a text and assigns a code number standing for a particular word sense to each word in that text. For each word, a lexicon is first checked to see what senses that word can take. Many words are unambiguous, but if more than one sense is possible for a given word, WSD techniques come into play, making use of the following types of information:

(1) The part of speech (POS) tag assigned by the CLAWS POS tagger (Garside et al. 1987). For example, if 'spring' is a verb, we know it must mean 'jump'. Wilks and Stevenson (1996) have shown that POS tagging greatly assists in the problem of word sense disambiguation.

(2) The general likelihood of a word taking a particular meaning, as found in certain frequency dictionaries. For example, if a corpus has 5,651 occurrences of 'bridge' in the sense of a bridge over a river, and only 194 occurrences of 'bridge' in the sense of a dental bridge, then the simplest technique is to assume that the more common sense of 'bridge over a river' is always the correct one. According to Allen (1995), this simple technique is about 70% accurate over English as a whole.

(3) Idiom lists are kept. If an entire idiomatic phrase is found in the text being analyzed, it is assumed that the idiomatic meaning of each word in the phrase is more likely than individual interpretations of the words.

(4) The domain of discourse can be an indicator. For example, if the topic of discussion is footwear, then 'boot' is unlikely to refer to the boot of a car.

(5) Special rules have been developed for the auxiliary verbs 'be' and 'have'.

(6) Collocations are pairs or groups of words that frequently appear in the same context. Are any collocates of the word, suggesting a particular interpretation, found in its immediate vicinity? We would, for example, prefer the dental sense of 'bridge' if the context contains collocates such as 'dentist' or 'cavity'. This technique is also called proximity disambiguation. The amount of text on either side of a word in which we look for collocates is called the window. One statistical measure of collocation strength is mutual information (Church and Hanks 1990).

SENSEVAL is an open evaluation exercise for WSD programs, first taking place in 1998 with WSD tasks for English, French, and Italian (Kilgarriff and Palmer 2000).

A corpus manually annotated with the correct sense of each word was used as a 'gold standard', against which the output of each of the programs was

concerned. The SENSEVAL systems could make use of the rich HECTOR dictionary rather than a simple lexicon. In HECTOR, for each homograph, there is a separate entry for each sense distinction, including fields for word sense definition, POS information, and examples of usage. This enabled the development of the following techniques:

(1) Collocation filters. These were used by various authors, the main variations being in (a) the window length and (b) whether to measure word–word, word–phrase, or noun–argument collocations. In order to calculate collocation statistics for verbs and the nouns they are associated with, shallow parsing is required beforehand.

(2) Manually created idiom extraction patterns, such as 'shake in \w* (shoes|boots|seat)' would cover any of the (variable form) idioms 'shake in your boots', 'shake in one's seat', etc.

(3) Matching dictionary example sentences with a window of input text. A score is given for the number of identical words occurring in the dictionary example and the text window. The sense with the highest scoring example is chosen. Similarly, one can count matches that involve semantically related words (such as words with the same Roget's thesaurus categories) in the matching score, rather than insisting on exact word matches. Various measures of similarity between test sequences and dictionary examples (glosses) have been suggested.

(4) Conceptual density. All possible senses of all content words in the input sentence are marked in a hierarchy such as Word-Net (Fellbaum 1998). The portion of the hierarchy with the greatest concentration of marked nodes (including one for the word being tested) will reflect the sense of the test word.

(5) Development of decision trees. According to the nature (word classes or individual words) of the neighbors of the test word, decisions are taken in an order designed to yield maximum information at each branch point, and to determine the sense of the test word in as few steps as possible. A similar technique for French grammatical words is described by Hug (2000).

(6) A number of systems use machine learning techniques for finding the combination of features (e.g. collocates, POS of words in the window) most likely to reveal word sense.

(7) Use of a thesaurus to overcome data sparseness. A recurring problem with WSD, compared with POS tagging, is that there are more word senses than syntactic categories, meaning a much larger amount of training data is required. Use of a thesaurus helps overcome this problem, as frequencies of word classes are studied rather than those of individual words.

In the next SENSEVAL evaluation, the WordNet hierarchy will be used rather than the HECTOR dictionary. Other interesting WSD methods described in the literature are as follows:

(1) Gale et al. (1992) used machine-readable texts and their translations, noting, for example, that the sense of 'drugs' which translates into French as 'mediacaments' collocates with 'prescription', 'patent', and 'generic', while the sense which translates as 'drogues' collocates with 'abuse', 'paraphernalia', and 'illicit'.

(2) Biber (1993) used the multivariate statistical technique of factor analysis to discover the four basic senses of the word 'right' in a corpus, according to their various collocates.

References

Allen, James. 1995. Natural language understanding. Redwood City, CA: Benjamin Cummings.

Biber, Douglas. 1993. Co-occurrence patterns among collocations: a tool for corpus-based lexical knowledge acquisition. Computational Linguistics 19(3). 531–8.

Church, Kenneth Ward, and Patrick Hanks. 1990. Word association norms, mutual information and lexicography. Computational Linguistics 16(1). 22–9.

Fellbaum, Christiane. 1998. WordNet, an electronic lexical database. Cambridge/London: MIT Press.

Gale, William, Kenneth Ward Church, and David Yarowsky. 1992. A method for disambiguating word senses in a large corpus. Computers and the Humanities 26(5–6). 415–39.

Garside, Roger, Geoffrey Leech, and Geoffrey Sampson (eds.) 1987. The computational analysis of English, a corpus-based approach. London: Longman.

Hug, Marc. 2000. Partial disambiguation for very ambiguous grammatical words. Journal of Quantitative Linguistics 7(3). 217–26.

Kilgarriff, Adam, and Martha Palmer (Guest Editors of the Special Issue on SENSEVAL). 2000. Computers and the Humanities 34. 127–34.

Thomas, Jenny, and Andrew Wilson. 1996. Methodologies for studying a corpus of doctor–patient interaction Using corpora for language research, ed. by Jenny Thomas and Mick Short, 92–109. Harlow: Longman.

Wilks, Yorick, and Mark Stevenson. 1996. The grammar of sense: is word-sense tagging much more than part-of-speech tagging? Technical Report CS-96-05, University of Sheffield.

MICHAEL P. OAKES

Working Memory

The term 'working memory' (WM) refers to the assumption that some form of temporary storage and manipulation of information within the time range of a few seconds is necessary for performing virtually all cognitive operations. Thus, understanding and analyzing a sentence, performing a mental arithmetic problem, or recalling a telephone number requires the activation of the working memory system.

A multicomponent WM system was proposed as an alternative conceptualization of the so-called short-term memory, considering the unitary system and experimental results, which led to its abandonment as a concept. The most commonly accepted WM model was proposed by Alan Baddeley and colleagues during the 1970s and consists of three main components: an attentional-controlling system (central executive) aided by two slave systems responsible for the temporary storage and manipulation of either visual/spatial material (visuospatial sketch pad) or acoustic/speech-based material (phonological loop) (Figure 1).

The central executive, the most important yet least well-understood component of the WM, is a limited-capacity processor that is responsible for providing the link between the slave systems and the so-called long-term memory. It is assumed to be an attentional control system and is responsible for online data storage and the selection, initiation, and termination of processing routines (e.g. encoding, storing, and retrieving). Recent findings point to distinct areas of the prefrontal cortex of the brain as neuronal substrate for the central executive (Figure 2), which controls the activity of the more posterior sensory areas that also contribute to WM. The prefrontal cortex appears to hold relevant information online as well as to perform complex processing functions.

The visuospatial sketch pad (or scratch pad) has a visual component, which is concerned with the analysis of color and shape, and a spatial component, which is concerned with location. Besides the important role of the prefrontal cortex for both components, the visual component has its neuronal correlates in bilateral occipitotemporal brain regions whereas the spatial component has its neuronal substrate in right occipitoparietal regions.

The phonological loop, which also comprises two components, is responsible for the storage and manipulation of sounds. The first component is a phonological store, which is capable of holding traces of acoustic or speech-based material for 1–2 seconds. The phonological store receives directly and unavoidably any information auditorily presented and stores it in terms of a sound-based code. It has its neuronal correlates in the left posterior parietal cortex and is also able to receive visually presented items, but these must first be converted into an articulatory form before gaining access to the store. The second component is an articulatory control process, which refreshes memory traces by subvocal rehearsal of phonological information. The neuronal substrate for subvocal rehearsal lies in left hemispheric speech areas, including Broca's area, the premotor cortex, and the supplementary motor area (SMA). Prevention of subvocal rehearsal results in very rapid forgetting.

Convincing evidence for the existence of such different components included in the WM model comes

Figure 1. Scheme of the multicomponent WM model after Baddeley (1986).
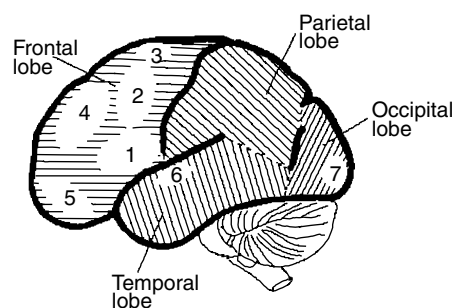
Figure 2. Left-hemispheric scheme of the brain lobes. In the frontal lobe, schematic positions of left-hemispheric brain areas are indicated, which are important concerning WM: Broca's area (1); premotor cortex (2); supplementary motor area (3); dorsolateral prefrontal cortex (4); orbital prefrontal cortex (5); primary auditory cortex (6); primary visual cortex (7).

from studies of brain-damaged patients with specific memory impairments. Some patients show selective deficits of auditory verbal WM, which was attributed either to selective impairment of the phonological store or of rehearsal processes. Other patients have a specific impairment of the visuospatial sketch pad or of the central executive.

WM and Language Processing

Concerning the relationship between WM and language processing, the phonological loop and the central executive are the most important components of the model. In fact, the visuospatial sketch pad might also play a role in some aspects of language processing, for example, in text comprehension in as much as mental imagery might be involved. The functional role of the phonological loop is probably to facilitate long-term phonological learning, as is needed to acquire both native and second language. Indeed, phonological learning is probably a primary determinant for successful first-language acquisition. It is necessary for the acquisition of new vocabulary and language comprehension as a whole. The function of the central executive is the inhibition of irrelevant information as well as activation and maintenance of information relevant during language comprehension. While listening to human speech, WM is used to hold the segments of sentences 'on-line' millisecond by millisecond. WM is necessary to carry forward, in real time, the subject of a sentence and associate it with verbs and objects in order to comprehend the sense and meaning of sentences. Some of the best evidence for WM playing a substantial role in language processing comes from studies that show substantial correlation between the measure of a person's verbal working memory capacity and the person's performance on language-understanding tasks. Therefore, subjects are divided into those with either high or low WM span, with the latter being more likely to be misled by inappropriate context. Generally, persons with a high working memory span seem to be better at language comprehension and syntactic analysis.

To test the influence of WM load on comprehension, some common paradigms in language research are used such as the comparison of simple subject–verb–object constructions (1a) with passive constructions (1b), which are harder to understand. Similarly, right-branching structures (2a) are easier to understand than center-embedded sentences (2b), since the main clause subject has to be kept in mind till the end of the embedded phrase. Furthermore, sentences with embedded relative clauses are used (3a) or constructions such

as in (3b), which are even more difficult to comprehend due to greater demands on WM. Subjects with low WM span have substantially more problems in understanding the more difficult sentences, and furthermore need more processing time.

(1) a. The dog chased the child.
   b. The child was chased by the dog.
(2) a. The child chased the dog that jumped from the table.
   b. The dog that jumped from the table chased the child.
(3) a. The dog that chased the child jumped from the table.
   b. The dog that the child chased jumped from the table.

WM is studied in humans mainly by the use of behavioral paradigms, most recently in conjunction with brain-imaging techniques. It has been shown that WM relies on cooperation among distributed areas of the brain, with the precise regions depending on whether tasks entail remembering objects, locations, or words. The prefrontal cortex is apparently working as the coordinator of the activity of these various regions.

References

Baddeley, Alan D. 1986. Working memory. Oxford: Oxford University Press.
———. 1999. Essentials of human memory. Hove: Psychology Press.
———. 2000. Short-term and working memory. The Oxford handbook of memory, ed. by Endel Tulving and Fergus I.M. Craik. New York: Oxford University Press.
Baddeley, Alan D., and Graham J. Hitch. 1974. Working memory. The psychology of learning and motivation, Vol. 8, ed. by Gordon A. Bower. New York: Academic Press.
Caplan, David, and Gloria S. Waters. 1999. Verbal working memory and sentence comprehension. Behavioral and Brain Sciences 22. 77–126.
Daneman, Meredyth, and Patricia A. Carpenter. 1980. Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior 19. 450–66.
Gathercole, Susan E., and Alan D. Baddeley. 1993. Working memory and language. Hove, Sussex: Erlbaum.
Gazzaniga, Michael S. (ed.) 2000. Cognitive neuroscience: a reader. Malden: Blackwell.
Just, Marcel, and Patricia A. Carpenter. 1992. A capacity theory of comprehension: individual differences in working memory. Psychological Review 99. 122–49.
Miyake, Akira, and Priti Shah (eds.) 1999. Models of working memory: mechanisms of active maintenance and executive control. New York: Cambridge University Press.
Roberts, Angela C., Trevor W. Robbins, and Lawrence Weiskrantz (eds.) 1998. The prefrontal cortex: executive and cognitive functions. New York: Oxford University Press.

SABINE WEISS

Writing Systems

Writing is distinguished from other forms of pictorial representation in that it relates directly to linguistic expression. Thus, a picture of a man feeding a dog, no matter how eloquent, is not writing. Writing as a means of representing linguistic expression probably first arose in Mesopotamia at the end of the fourth millennium BC, although some recent discoveries suggest that the earliest writing in Egypt might be contemporaneous or even earlier.

From a linguistic point of view, writing systems can be classified in terms of the kinds of linguistic elements that form the basic units of the system, with a basic division into logographic, syllabic, and alphabetic types, although there are also intermediate types (such as alphasyllabic) and writing systems that combine more than one type. The scale logographic–syllabic–alphabetic also corresponds approximately to the historical development of writing systems.

In a logographic system, there is a separate symbol for each morpheme of the language in question. The closest modern writing system to a pure logographic system is that of Chinese. In (Mandarin) Chinese, the words for 'lake' and 'pot' are both pronounced alike, *hú*, but have written forms that are completely different, as in examples (a) and (b), respectively. Although the term logographic might suggest one symbol per word, this is not strictly accurate, since a word consisting of more than one morpheme, such as *wǒ-men* 'we', literally I-PLURAL, is written with two symbols, one for each morpheme, as in (c).

湖　　壺　　我们
(a)　(b)　(c)

In a syllabic system, there is a separate symbol for each syllable. Japanese uses two syllabaries (sets of syllabic symbols), illustrated here with the hiragana syllabary. The symbols in (d), (e), and (f) are, respectively, the syllables [ta], [to], and [no]. Note that there is no similarity between the symbols for [ta] and [to] corresponding to the shared initial [t], and no similarity between the symbols for [to] and [no] corresponding to the shared vowel [o].

た　　と　　の
(d)　(e)　(f)

In an alphabetic system, there is a separate symbol for each phoneme of the language. The Latin, Greek, and Cyrillic alphabets (the last used, for instance, for Russian) are all at least in principle alphabetic writing systems, so that in regularly spelled English words like *ran, run, bun, but,* each orthographic symbol (*a, b, r, n, t, u*) corresponds to a single phoneme. The earliest alphabetic writing systems departed from this pattern in an interesting and systematic way: they wrote only consonants. This consonantal writing system survives, with certain modifications, in the modern Arabic and Hebrew writing systems. The Arabic word as written in (g) represents only the consonants [k], [t], and [b], in that order (although read from right to left in Arabic). The word would most probably be read as *kataba* 'he wrote', although it could in principle also be read as *kutiba* 'it was written'. It is possible to add diacritics to indicate the vowels, as in (h) for *kataba* and (i) for *kutiba*, but it is not usual to do so, since the correct interpretation can usually be derived from the context by someone with native-like fluency in reading the language. The first alphabet to use both consonant and vowel symbols consistently was the Greek alphabet.

كُتِبَ　كَتَبَ　كتب
(g)　　(h)　　(i)

An interesting halfway stage between alphabetic and syllabic writing is found in the so-called alphasyllabic writing system, found in most of the indigenous scripts of South Asia (e.g. Devanagari, Bengali, Tamil), South-East Asia (e.g. Burmese, Thai), and Ethiopia and Eritrea (e.g. Amharic). Here, the basic symbols of the writing system represent consonants, but modifications are added to indicate vowels, thus giving rise to complex symbols that represent syllables. Examples (j), (k), (l), and (m) are from the Devanagari script, used for writing modern Hindi, and represent, respectively, [ke], [pe], [ku], and [pu]. The vowel [e] is represented by a kind of hook above the consonant symbol, and the vowel [u] by a kind of hook below.

के　पे　कु　पु
(j)　(k)　(l)　(m)

Occasionally, one finds orthographic representation of distinctive features, i.e. the individual phonetic parameters that make up a phoneme, although only in particular parts of the system. In Czech, for instance,

which uses the Latin alphabet, palatalization is indicated by means of a so-called hook placed above the nonpalatalized correspondent, so that alongside *s*, *z*, and *c* (the last like the initial consonant of *tsetse*), one has *š* (like the initial consonant of *ship*), *ž* (like the final consonant of *rouge*), and *č* (like the initial consonant of *chip*).

Few writing systems correspond exactly to one or other of the ideal types. The Chinese writing system, although basically logographic, makes some concessions to pronunciation, so that like-sounding words sometimes incorporate an identical element. In Mandarin Chinese, the pronunciation *huáng* corresponds to the three meanings (among others) 'emperor', 'locust', and 'afraid', which use the symbols shown in (n), (o), and (p), respectively. It will be seen that the symbols for 'locust' and 'afraid' include as their right-hand component the symbol for 'emperor'. In addition, symbols can be used in writing foreign words purely for their phonetic value; thus, the Chinese for Berlin is *bólín*, written as in (q), where the two symbols normally represent, respectively, 'cypress' and 'grove'. This is essentially the same as the rebus principle often used in children's picture writing puzzles, where for instance a picture of a bee might represent the like-sounded verb 'be'.

皇　　蝗　　惶　　柏林
(n)　　(o)　　(p)　　(q)

Although vowels are not universally written in Arabic, long vowels nearly always are, a departure away from a purely consonantal writing system, so that the word *kaatib* 'secretary', as in (r), does differ in writing from item (g).

كاتب
(r)

Probably the most complex mixed script in use today is that of Japanese, which uses (simplifying somewhat) logographic symbols of Chinese origin for roots (e.g. *hito* 'man', *mi-* 'see'), syllabic symbols from the hiragana syllabary for particles and suffixes (e.g. nominative *ga*, accusative *o*, past tense *-ta*), and syllabic symbols from the katakana syllabary for loanwords from other languages (e.g. *doa* 'door'), as illustrated in (s).

人	が	ドア	を	見た。
hito	ga	do-a	o	mi-ta.
man	NOMINATIVE	door	ACCUSATIVE	see-PAST

'The man saw the door.'
(s)

English is notorious for its departures from the regular principles of alphabetic writing, so that we find words with the same pronunciation but different spellings (e.g. *bear, bare*), words with the same spelling but different pronunciations (e.g. *lead* as in *I will lead you to the restaurant*; *lead* as in *lead piping*), as well as words whose spelling seems to bear no relation to their pronunciation (e.g. *eye*). As Bernard Shaw once observed, from the spelling of the [f] of *enough*, the [ɪ] of *women*, and the [ʃ] of *nation*, one might deduce that *fish* ought to be spelled *ghoti*! This contrasts with some other languages using the Latin alphabet, where the phonemic principle is strictly observed, as in Finnish, or at least more strictly observed, as in Spanish.

It is interesting to ask why such discrepancies between spelling and pronunciation exist. Occasionally, they simply reflect random decisions by scribes that happen to have gained acceptance into the language, as in the spelling *delight*, where the *gh* has no justification, either synchronically or diachronically. In many instances, however, irregular spellings reflect earlier pronunciations, i.e. the spelling was originally justified in terms of the pronunciation at that time, and in the meantime the pronunciation has changed but the spelling has remained conservative. The *gh* of such words as *light* and *bright* was pronounced like the *ch* sound in *loch* until medieval times in English, and indeed it is still so pronounced in some Scottish dialects; similarly, until medieval times the initial consonants of *knee* and *gnaw* were pronounced. Where another language has or had prestige in a community, loan words may retain, in whole or in part, the spelling of the original language, which accounts for the initial *p* in *psychology* (from Greek, the source of much English scientific terminology).

One recurrent deviation that is found from phonemic writing in many languages of the world is the application of the so-called morphophonemic principle, whereby the same morpheme is written in the same way even when its pronunciation changes. In English, for instance, the regular plural suffix is written *-s* even though it is sometimes pronounced [s] (after voiceless consonants, e.g. *cats*), sometimes [z] (elsewhere, e.g. *dogs*, *zebras*). Likewise, unstressed vowels in English are often reduced in pronunciation to schwa, but the same spelling is retained as when the vowel is stressed; compare the spelling of both stem vowels in *torrent*, stressed on the first syllable, and *torrential*, stressed on the second syllable. Another example of morphophonemic spelling in English is the invariant spelling of the root in the adjective *divine* and the derived noun *divinity*, even though the stressed vowel is pronounced [ai] in the former, but [ɪ] in the latter.

It is often difficult to tell whether certain spellings are really historical or morphophonemic, since morphophonemic alternations often reflect sound changes that have differentially affected an originally identical sound in different environments. Thus, the spelling of the Russian word for 'wave' has an *o* in the first unstressed syllable in the singular, as in (t) (corresponding to *volna* in Latin script), even though this unstressed vowel is pronounced [a]. This can be explained both historically and morphophonemically. Historically, because this was an earlier pronunciation, still preserved in northern dialects, whereas other dialects have merged earlier unstressed [o] and [a]. Morphophonemically, because this vowel is stressed and pronounced [o] in the plural, as in (u) (corresponding to *volny* in Latin script).

волна	волны
(t)	(u)

The development of writing must be seen as one of the greatest achievements of humanity, permitting, even before the advent of audio and video recording, the preservation of texts for future generations and their transmission to distant locations. The decipherment of such ancient writing systems as Egyptian hieroglyphs in the nineteenth century and Mayan glyphs in the twentieth century is not only an intellectual achievement on the part of the decoders, but has also provided us with new insights into societies of the past.

References

Daniels, Peter T., and William Bright (eds.) 1996. The world's writing systems. New York: Oxford University Press.

Hooker, J.T. 1990. Reading the past: ancient writing from cuneiform to the alphabet. Berkeley: University of California Press and London: British Museum.

Pope, Maurice. 1999. The story of decipherment: from Egyptian hieroglyphs to Maya script, revised edition. London: Thames and Hudson.

Robinson, Andrew. 1995. The story of writing: alphabets, hieroglyphs and pictograms. London: Thames and Hudson.

Sampson, Geoffrey. 1985. Writing systems: a linguistic introduction. London: Hutchinson.

BERNARD COMRIE

Y

Yémba and the Grassfields Bantu Languages

The Grassfields Bantu Languages

The Grassfields languages form the principal subgroup of Southern Bantoid, within the Niger-Congo phylum that dominates sub-Saharan Africa. Some 65 Grassfields languages are spoken in the mountainous plateau of the western and northwestern provinces of Cameroon (see Figure 1), the region that is thought to be the origin of the original Bantu expansion. These languages exhibit predominantly subject–verb–object word order, a noun class system (typically prefixes), noun class agreement between words in the noun phrase, and complex tone systems. Many of these features are illustrated below for Yémba. The segmental inventories of Grassfields languages typically include (prenasalized) stops, fricatives, affricates, liquids, nasals and glides with labial, labiodental, alveolar, palatal, velar and glottal places of articulation. Some languages also exhibit labiovelars and implosives.

Key distinguishing features of Grassfields languages with respect to narrow Bantu are the simplified verbal morphology and the smaller number of noun classes, the loss of final vowels from Proto-Bantu roots, the 'floating' tones that interact in highly complex ways with their surroundings, and the addition of a high central vowel to the five-vowel system of Proto-Bantu.

Within Grassfields, three main subfamilies are distinguished: Mbam-Nkam (34 languages), Ring (16 languages), and Momo (eight languages). Within Mbam-Nkam four subgroups are distinguished, the largest being the Bamiléké languages (the name 'Bamileke' is a corruption of mbə ləkʉɔ, literally *people of the ravine*).

Yémba

Yémba is a Bamiléké language spoken by over 300,000 people in the region around the town of Dschang [tʃaŋ], and the language is most widely known by this latter name. About ten dialects have been identified, the most significant being those of Bafou to the north and Foréké-Dschang to the south. A Roman-based orthography was established in the early 1930s and literacy was promoted first at a school in the paramount chief's compound, and later in mission and government schools. Around the time of independence in 1960, education in local languages was forceably halted because it was viewed as tribalism and an obstacle to nation building. By 1980 the political climate had changed and the development of local languages was linked to the construction of a new noncolonial pan-African identity. A national alphabet based on the International Phonetic Alphabet (IPA) and derived from the Africa script was introduced, and the Yémba orthography was duly modified. The following text illustrates the orthography, where acute accent marks high tone and macron marks mid-tone.

> Kaŋ pɔ́ mbhū̄ é lelá' ŋ́gɔ̄ mésō, ḿbú ńzīŋɛ́ ta' enɔ. Pɔ́ lelá' ńnāŋ tɛ eshū̄' amɔ̄' álī'í, ḿbɛ́ á ápa, ńdɔk ŋgū̄ɔ́ á ŋ́kā' ŋiŋ njʉ́ʉ́ a apumā. Pɔ́ le gē é tō á ḿba, ŋ́gɔ́ɔ mbɔŋ. Pɔ́ le gū̄ɔ́ tɛ ŋkó éwú, kaŋ á le mɛ́ mbhū̄ ŋ́gɛː «Esó, pá' meŋ ŋkʉ́ɔ́k ḿbiŋ nzéŋ ńzāŋné lā, meŋ ɛ̄ kɔ́'á áthū̄, é kāp, o gɔ́ á ésī ńnɔŋ mbétʒ́ ńnéŋ á ńtē á ápa».

Yémba lacks the labiovelars and implosives found in some of the other Grassfields languages, and it has augmented the Proto-Grassfields six-vowel system with a series of mid-vowels. Diachronic processes,

Cameroon Language Map
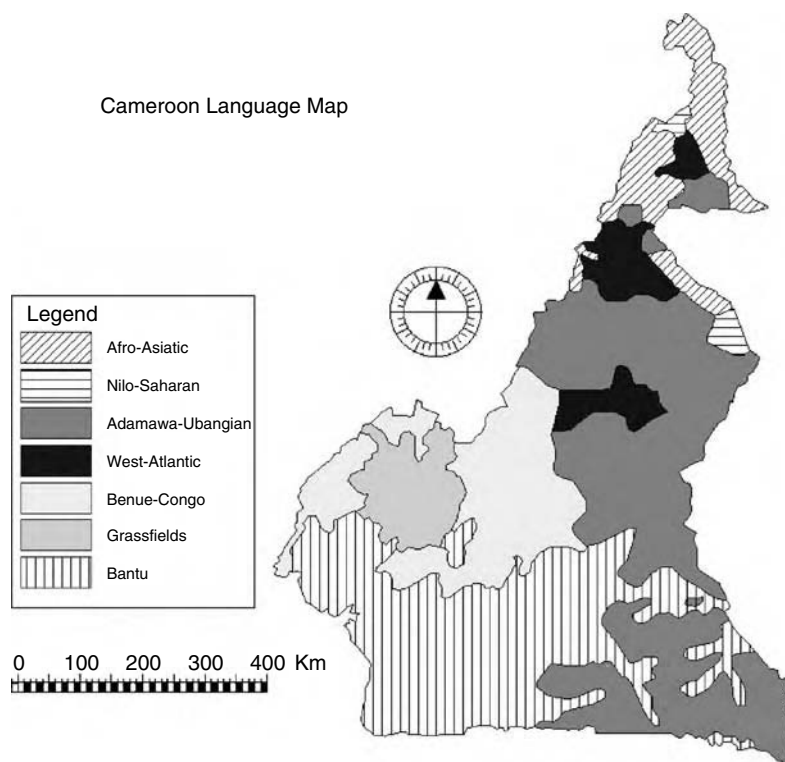
Figure 1. Cameroon language map showing principal families.

such as the loss of final consonants and the devoicing of high vowels, are still manifested synchronically in some complex morphophonological processes. Notable consequences of these processes are highly complex tonal patterns, and heavily aspirated syllables. This aspiration is most striking after voiced obstruents (e.g. /ghíɛ́ / [ɣiɛ́] *fly!*).

Like the other Grassfields languages, Yémba exhibits lexical and grammatical tone. The following nouns have a lɔ́- class 5 prefix, followed by the tə̌ŋ syllable that carries H, HL, LH, and L tones (from Proto-Bantu HH, HL, LH, and LL): lə̀tɔ̌ŋ [-ʹ] *feather*, lə̀t⁺ɔ́ŋ [-ˉ] *reading*, lə̀tɔ̀ŋ° [--] *navel*, and lə̀tɔ̀ŋ [-ˌ] *finishing*. These examples demonstrate that tone is a lexical property of words. However, tone in Yémba also has a grammatical function. The following examples illustrate how tonal distinctions alone can convey tense distinctions. The words used in the examples are: ə̀-fɔ̀ *CL1-chief*, kə̀mtè *bury*, and mə̀m-bh⁺ú *CL10-dogs* (ə̀- and mə̀m- are class 1 and 10 noun class prefixes). The words ɔ́ and ɔ̀ are known as concord markers, serving no purpose other than to link the subject and verb.

(1) ə̀fɔ̀ ɔ́ ⁺kə̀mtè ⁺mə̀mbhú [-l-ˉl--l-] 'the chief buried dogs (immediate past)'
(2) ə̀fɔ̀ ɔ̀ kə̀mtè mə̀mbh⁺ú [-l--l---l-] 'the chief buries dogs (simple present)'
(3) ə̀fɔ̀ ɔ̀ ⁺kə̀mtè ⁺mə̀mbhú [-l--l--l-] 'the chief will bury dogs (immediate future)'

The investigation of Grassfields tone systems has stimulated important theoretical advances in autosegmental phonology.

References

Bird, S. Orthography and identity in Cameroon. Written Language and Literacy 4. 131–62.

———. 1999. Dschang syllable structure. ed. by H. van der Hulst and N. Ritter, The syllable: views and facts, Studies in Generative Grammar, 447–76. Berlin: Mouton de Gruyter.

Bird, S., and Tadadjeu, M. 1997. Petit Dictionnaire Yémba–Français (Dschang–French Dictionary). Cameroon: ANACLAC.

Hyman, L.M. 1972. A phonological study of Fe?fe?-Bamileke. Studies in African Linguistics (Supplement 4).

———. 1980. Relative time reference in the Bamileke tense system. Studies in African Linguistics 11. 227–37.

———. 1981. Noni grammatical structure. Occasional papers in linguistics, Vol. 9. University of Southern California.

———.1985. Word domains and downstep in Bamileke-Dschang. Phonology Yearbook 2. 45–83.

Hyman, L.M., and M. Tadadjeu. 1976. Floating tones in Mbam-Nkam. Studies in Bantu tonology, ed. by L.M. Hyman, 57–112. University of Southern California. Occasional Papers in Linguistics, Vol. 3.

Voorhoeve, J. 1971. Tonology of the Bamileke noun. Journal of African Languages 10. 44–53.

Watters, J.R., and Leroy, J. 1989. Southern Bantoid. The Niger-Congo languages, ed. by J. Bendor-Samuel, 430–49. Lanham, MD: University Press of America.

STEVEN BIRD

Yiddish

Yiddish is the historical language of the North European 'Ashkenazic' Jews and their descendants worldwide (Ashkenaz denoted an Iranian people in the Bible, which in the Middle Ages was used to refer to Slavic, and then, by about 1100, to Germans and German Jews). Since the 1600s, the language has been called *jidiš* 'Jewish' (earlier native names are unknown), and more recently *mame-lošn* ('mother tongue') and *žargon* ('jargon', often contemptuous in languages other than Slavic).

Yiddish is traditionally regarded as a derivative of High German, which first formed in southern Germany c. 1000 CE. In the thirteenth century, Yiddish speakers began migrating to Poland; dialects spoken there and in Belarus', Ukraine, Russia, the Baltic lands, and Romania are called 'Eastern Yiddish', which is used by most contemporary speakers of the language. Speakers in Holland, Hungary, the German-speaking and Bohemian-Moravian lands, however, use what is known as 'Western Yiddish'. Although the two spoken Yiddishes were always distinct, the Eastern Yiddish literary language was largely based on Western norms until the early nineteenth century. A new view holds that Eastern and eastern Western Yiddish (in the originally bilingual eastern German and Slavic lands) are instead a multilayered Slavic language with a predominantly German lexicon formed between the ninth and twelfth centuries, while western Western Yiddish (best called 'Judeo-German') is Germanic, having been formed by French Jews in the Rhineland between the tenth and late fourteenth centuries (and now extinct). Both Romance and Slavic Jewish settlements in Germany (the latter of local and Balkan stock) date to the ninth to tenth centuries. Hence, Yiddish is the only living Indo-European language whose genetic assignment is still being debated.

The mass migration of speakers to Palestine and the Americas, beginning in the 1880s, resulted in most migrants eventually 'losing' their language; they stopped speaking their first language and switched to the predominant language of their new homeland. Language shift from Yiddish to Russian was also rapid in the Soviet Union, without outward migration. Nevertheless, in 1939, Yiddish speakers probably numbered ten million worldwide. Accelerated by the Nazi German extermination of the European Jews, the number of speakers today, found primarily in Israel and the Americas, has dropped to around one million. Although Yiddish never enjoyed official status except in the 'Jewish Autonomous Region' of Eastern Siberia (Russia) formed after the Revolution, and in Soviet Belarus' and Ukraine (until the early 1930s), it has been the object of organized standardization efforts in Europe and America since 1908.

In addition to being Indo-European, Yiddish is a member of a genetically mixed 'family' of some two dozen languages adapted by Jews from non-Jewish linguistic stock, such as Arabic, Berber, Chinese, Greek, Iranian, Romance, and Slavic, and it therefore owes its *raison d'être* to common processes of formation and (Hebrew) enrichment. Of the Jewish languages, Yiddish has by far the largest number of speakers, and the richest literature. Yiddish was taught in the early 1700s in German universities, and since the twentieth century has stood at the cutting edge of modern linguistic theory, particularly since top theoreticians such as Edward Sapir, Roman Jakobson, Max and Uriel Weinreich, and Edward Stankiewicz made it their occasional or sole area of interest. Yiddish was the first Jewish and Soviet language to be the subject of a linguistic atlas (see Vilenkin 1931, and now Herzog et al. 1992–2000).

Taking Western and Eastern Yiddish as a whole, Yiddish enjoys the broadest expanse of any European language, with the exception of Russian and perhaps Romani (Gypsy); in the sixteenth century Yiddish was spoken from Holland to Ukraine, and from the Baltic Sea to Northern Italy and in Ottoman Palestine. While Yiddish was obsolescent in its German homeland and western Hungary from about 1750 (being displaced by standard German, after passing optionally through a stage of slightly Yiddishized German until the early twentieth century), it reached Russia and America in the late eighteenth to early nineteenth century, and by the late nineteenth century had arrived in the other inhabited continents, including remote Harbin (Manchuria) and Shanghai. In an unusual sociolinguistic development, Yiddish is now obsolescent for the most part except among ultra-orthodox Jews.

Yiddish is largely written phonetically in the Hebrew (Aramaic) alphabet, although Hebraisms and Aramaisms are spelled etymologically (except in Soviet Yiddish, which uses an entirely phonetic spelling). Eastern Yiddish generally comprises approximately 75% Germanisms, 15% Hebraisms, and 10% Slavisms; in German-speaking milieus, Yiddish often acquires a much higher Hebrew component so as not to be easily understood by German speakers (some of

these Hebraisms entered German beginning with the 1400s, then—through German—into other European slang registers). There are also small unique Romance, Greek, Turkic, and Iranian components.

The origin of Yiddish has recently become a key topic in Yiddish linguistic discussion. In the 'traditional' model of Yiddish genesis, proposed by Max Weinreich (1973), the language's unique Romanisms suggest the founders of Yiddish came from France and Italy. Weinreich envisioned Yiddish born in the Rhineland and Regensburg, then largely reshaped in Bavaria-Franconia, prior to spreading into monolingual Slavic lands. Seven hundred years of contact with Slavic, beginning with Sorbian and Polabian in the ninth century and culminating with Russian in the eighteenth century, rendered parts of Western and all of its Eastern Yiddish offspring heavily Slavicized. Examples of Slavic features are the possessive pronouns inflected only for number; the leveling out/addition of morphophonemic alternations in Germanisms; the distribution (and often indeclinability) of the reflexive pronoun *zix*; two degrees of diminutive; the double negative; analytic expression of comparative and superlative adjectives; recalibration of German verbal prefixes; psycho-ostensive expressions; the resumptive pronoun; no verb-final position in subordinate clauses; exclusive use of the present perfect to express past time; initial *s* and *x* (in non-German components); the absence of [ç]; and final voiced consonants (the last feature was probably original, and not restored under Slavic influence). While some of these features are found in German dialects (often spoken in former Slavic areas), Eastern Yiddish almost never has 'German' features unknown to Slavic.

Since the late 1980s, the obvious similarity of Yiddish to Bavarian dialects and the lack of features from southwestern German led linguists to favor a (south)eastern birthplace. The Danube rather than the Rhine venue is further supported by the likelihood that the few unique old Romance elements all derive from Balkan Romance languages, Friulan, Rhaeto-Romance, or Italian and hardly ever from French, except in Germanic Western Yiddish. The origin, language, and fate of the Jews in fourth-century Roman Rhineland are unknown.

Between 1991 and 1993, while accepting an eastern German venue for the birth of Yiddish, Wexler proposed that Sorbian Jews in Eastern Germany—comprising mainly local and Balkan proselytes and some ethnic Jews—'relexified' their West Slavic language to High German by the twelfth century. Relexification means that the Jews retained Sorbian grammar, phonology, and phonotactics while replacing most of the original lexicon with German words—whose meanings were assigned by the original Slavic vocabulary they replaced (see Horvath and Wexler 1997). In the kind of vocabulary replacement known as relexification, the substratal Slavic semantic and derivational parameters are kept, so that one can predict with considerable accuracy which superstratal German components can be accepted—unlike straightforward bilingual interference, which is always unpredictable and usually involves vocabulary exclusively. Parts of the German lexicon incompatible with Sorbian are blocked in Yiddish, to be replaced by genuine and many newly coined Hebraisms, unrelexified Slavisms, or other acceptable Germanisms. For example, German uses a common root for *Gewitter* ('storm') and *Wetter* ('weather'). Wexler expects relexifiers to accept at best one Germanism because Upper Sorbian denotes the two differently; see *hrimanje* and *wjedro*, respectively. Not surprisingly, Yiddish has only *veter* 'weather'. The use of separate words in Slavic makes German *Sturm* 'storm' possible for Yiddish (as *šturem*). Only relexification can account for the mass of Hebraisms as well as the very reduced German lexicon in Yiddish (and almost total impoverishment of German synonyms), compared with Central and Bavarian German dialects that conceivably were the relexifier sources for Yiddish. Following the relexification process that created it, Yiddish came to borrow additional German, Hebrew, and Slavic lexicon wihout reference to Slavic grammar.

Yiddish speakers have an enormous Hebrew-Aramaic lexicon (in all semantic domains) and a morphology used mainly with Semitisms, unmatched by any other Jewish language. This is because the blockage of so much German vocabulary obliged Yiddish speakers to invent many Hebrew forms and/or meanings as replacements, such as Yiddish *nadn* ('dowry') and *xmime* ('extreme heat'—an invention of Yiddish speakers, based on Slavic lexicons which, unlike German, have a word for 'extreme heat' that differs from the usual word for 'hot') vs. Hebrew *nādān* ('sheath') and *ħam* 'hot'. The Yiddish pronunciation of Hebrew predates the emergence of Yiddish, but the origins remain unclear. Yiddish Bibles, first attested in Germany in the 1400s, reveal many atypical Germanisms, and slavishly follow Hebrew syntax and derivational patterns; hence, they are really 'relexifications' of Biblical Hebrew into Yiddish lexicon, i.e. they utilize the original Old Hebrew grammar but replace almost all of the Old Hebrew lexicon by German-origin words. The latter are often used in ways that are unintelligible to German speakers. 'Yiddish' Bible language is best defined as a lexically deviant variant of Old Hebrew.

A Slavic affiliation for Yiddish is supported by its Slavic-type syntax, phonology, and phonotactics, as

well as the role of the derivational morphology; only the lexicon (including the inflectional and most of the derivational morphology) is predominantly German. A specifically Sorbian substratum for Yiddish is supported by the shape and geography of some unrelexified Yiddish Slavisms and the striking parallels in the distribution of Hebraisms in Yiddish vs. Germanisms in Sorbian, which points to a common speech community. Hence, Eastern Yiddish is probably an unusual dialect of Sorbian rather than a Slavicized form of German.

There is agreement that Yiddish speakers reached the Belarusian and Ukrainian lands in the 1400s, where they encountered Slavic-speaking Jews. The 'Yiddish-is-Germanic' school believes that the latter, few in number, shifted to Yiddish. In the 'Yiddish-is-Slavic' model, the local East Slavic Jews allegedly carried out the process of relexification a second time, as is borne out by the existence of uniquely East Slavic grammatical features in Yiddish, such as gender correspondences that differ from those found in German, and the original use of the German plural suffix *-(e)n* as a dual (significantly, the Yiddish dual matches the distribution of the Belarusian-Ukrainian 'pseudo-dual' ending used after 2–3–4; the dual is a category that has never been acquired through language contact). The attraction to Yiddish by East Slavic-speaking Jews attests to their high regard for Ashkenazic culture and desire to differentiate themselves more sharply from the coterritorial non-Jews; their ability to reform imported Sorbian Yiddish attests to their numerical superiority over the Ashkenazic immigrants. Indigenous East Slavic Jews could only be descended largely from the Turko-Iranian converts to Judaism in the Khazar Empire. A Khazar component in Yiddish is also suggested by a few unique Iranian and Turkic elements: e.g. Yiddish *šabaš* 'tip given to musicians at a wedding by guests who join in the dancing'; *šibeš* 'small coin; trifle' stems from Persian *šābāš* (in Slavic and other European slang vocabularies, the word is formally and/or semantically distorted); *nar*, plural *naronim* 'fool', derives from Persian *ner* 'man' rather than from German *Narr* 'fool' ~ Old Ukrainian *nar'-ci*, *nor'ci* 'Ossetian epic heroes' (Yiddish *-on-* from Old Hebrew and Old Ukrainian *-ci* both express an agentive noun); *pejsex* 'Passover' comes from Hebrew, but its use as a male name for children born during the holiday mirrors Turkic practice (the word as a name is first attested in tenth-century Khazar Hebrew); and a periphrastic conjugation that combines Hebrew masculine singular participles, now indeclinable, with a Yiddish auxiliary (e.g. Hebrew *bōdeq* 'inspecting' became Yiddish *bojdek zajn* 'inspect', with *zajn* 'be' from German). A Turko-Iranian origin for this conjugation is suggested by its geography within Yiddish (it is much less productive in Western Yiddish and altogether unknown in German slang), and by its use in Judeo-Belarusian (attested in the seventeenth century) and in the Turkic and Iranian languages spoken by Muslims (for Arabic loans).

Many scholars speak of a Western–Eastern Yiddish dialect continuum, with Western providing the input for Eastern Yiddish (Herzog et al. 2000). Yet it is also commonly assumed that Eastern Yiddish dialects developed in situ, rather than being imported from Western Yiddish 'readymade'. A multilayered West–East Slavic affiliation for Yiddish requires exploring, in future, to what extent Eastern Yiddish dialects match differences among the dialects of coterritorial Slavic languages.

Regardless of its origin, Yiddish holds an abiding interest for students of Jewish linguistics and history, bilingualism, minority ethnolects, creole linguistics, German, Slavic, unspoken Ashkenazic Hebrew, Modern Hebrew, and Esperanto (the last three are demonstrably Slavic Yiddish—the first two relexified to Old Hebrew and the third to Latinoid roots). This is because Yiddish provides clues to Jewish ethnogenesis and migrational patterns that are not recoverable from the notoriously sparse historical and archeological evidence; Yiddish is an ideal laboratory for the study of relexification, since the dual processes of relexification in that language are considerably older than the relexification found in numerous creole languages (e.g. Haitian Creole); finally, repeated exposure over a millennium to the same languages (German, Slavic, Hebrew) makes Yiddish a unique laboratory for students of bilingual interference.

References

Birnbaum, Salomon A. 1979. Yiddish: a survey and a grammar. Toronto-Buffalo: University of Toronto Press; includes a comprehensive bibliography.

Herzog, Marvin, et al. (eds.) 1992–2000. Language and culture atlas of Ashkenazic Jewry, 3 vols. Tübingen-New York: Max Niemeyer and YIVO; ongoing publication.

Horvath, Julia, and Paul Wexler (eds.) 1997. Relexification in Creole and non-Creole languages (with special attention to Haitian Creole, Modern Hebrew, Romani and Rumanian). Wiesbaden: Otto Harrassowitz.

Mark, Yudel, et al. (eds.) 1961–1980. Grojser verterbux fun der jidišer šprax, 4 vols. New York-Jerusalem: Yiddish Dictionary Committee; incomplete.

Šapiro, Mojsej Al'terovič, et al. (eds.) 1984. Russko-evrejskij slovar'. Moscow: Russkij jazyk.

Vilenkin, Lazar. 1931. Jidišer špraxatlas fun sovetn farband afn grunt fun di dialektologiše materjaln, vos zajnen cuno-jfgezamlt gevorn durx der špraxkomisje fun jidišn sekter fun der vajsrusišer akademje unter M. Vejngers onfirung. Minsk: Vajsrusiše visnšaft-akademje, jidišer sekter.

Weinreich, Max. 1973. Gešixte fun der jidišer šprax 4 vols. New York: YIVO; partial English translation, The history of the Yiddish language. Chicago: University of Chicago Press, 1980.

Weinreich, Uriel. 1968. Modern English–Yiddish Yiddish–English dictionary. New York: YIVO-McGraw-Hill.

Wexler, Paul. 1990. References and bibliography of Yiddish linguistics, 1979–1988. Studies in Yiddish linguistics, ed. by Paul Wexler. Tübingen: Max Niemeyer.

———. 1991. Yiddish—the fifteenth Slavic language. International Journal of the Sociology of Language 91. (focus issue). 1–150, 215–25.

———. 1992. The Balkan substratum of Yiddish (a reassessment of the unique Romance and Greek components). Wiesbaden: Otto Harrassowitz.

———. 1993. The Ashkenazic Jews. A Slavo-Turkic people in search of a Jewish identity. Columbus, Ohio: Slavica.

———. 2002. Two-tiered relexification in Yiddish. Jews, Sorbs, Khazars and the Kiev-Polessian dialect. Berlin-New York: Mouton de Gruyter.

PAUL WEXLER

See also **German; Hebrew: Biblical; Hebrew: Modern; Old Church Slavonic**

Yoruba and Yoruboid Languages

Yoruboid languages (YRB)—Yoruba, Igala, and Itsɛkiri—are classified as the Defoid group, a branch of Benue-Congo, which is a subgroup of Niger-Congo languages. Yoruba, the most widely spoken of the group, has over 20 million speakers in southwestern Nigeria, southeastern Benin, and central and northern Togo. In Sierra Leone, the influence of Yoruba is evident in Krio loanwords and personal names. Outside West Africa, Yoruba is used in religious contexts in Brazil and Cuba, as well as in cultural revitalization movements such as the ɔyɔtũɟi village, a Black Nationalist community in South Carolina. Yoruba has over 20 distinct dialects. In Nigeria, Standard Yoruba is the literary usage and is taught in schools up to the university level. It is also spoken by younger generations as a second or third language because of the implementation of the 1981 National Policy on Education, which requires high school students to learn a major Nigerian language in addition to their mother tongue. Furthermore, Yoruba is the official language in southwestern Nigeria and is used for government notices, radio, television, and newspapers. Igala is spoken by approximately 800,000 people in central Nigeria and Itsɛkiri is spoken by about 500,000 people in midwestern Nigeria. Igala and Itsɛkiri are taught in primary schools and are used in radio and television broadcasting. The distribution of Yoruboid languages is shown in Map 1.

Phonology (Sound Patterns)

The sound inventory of YRB includes the consonants shown in Table 1. In the Stop series, voiced (sounds produced with the vibration of the vocal cords) and
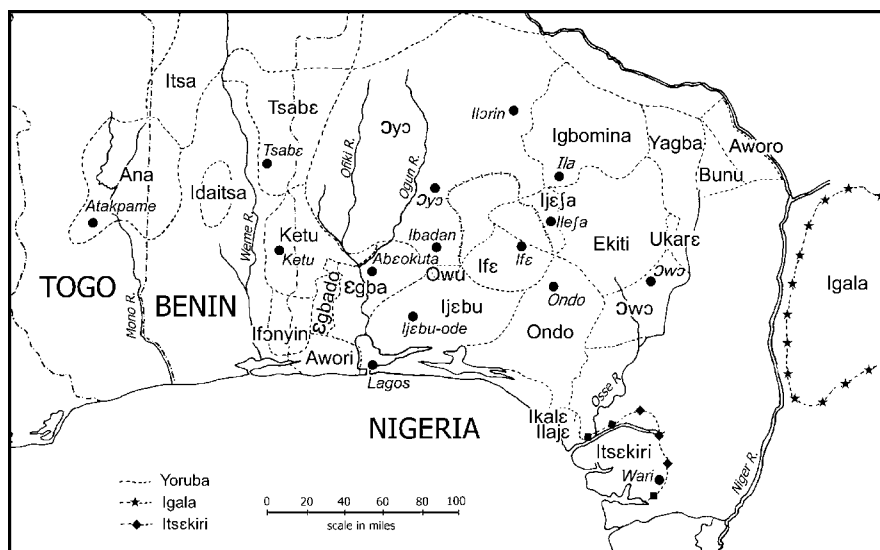
Map 1. Yoruba and Yoruboid languages.

unvoiced (sounds produced without vocal cord vibration) consonants are attested. However, /b/ and /ɟ/ have no unvoiced counterparts. The velar stops—kʷ and gʷ—are found in Igala, Itsɛkiri, and southeastern dialects of Yoruba such as Ondo and, but they are realized as /k/ and /g/ in Standard Yoruba and northwestern dialects. Fricatives (sounds produced with audible friction) are generally unvoiced. The only voiced fricative, /ɣ/, is attested in Itsɛkiri and southeastern Yoruba dialects, and it is the equivalent of /w/ in Igala and northwestern Yoruba dialects. Yoruba /s/ may correspond to Igala [r], [l], or [h], depending on the context. An /s/ occurring before an oral vowel (vowel pronounced through the mouth) is an [r] in Igala (Yoruba /ɛsɛ̀/ is realized as /ɛ́rɛ̀/ 'foot'); when /s/ precedes a nasal vowel (vowel pronounced by allowing air to pass out through the nose), it is realized as [l] (Yoruba /ɛ̀sɔ̃́/ is Igala /ɛ̀lá/ 'nine'); an /s/ occurring before /i/ is an [h] (Yoruba /ɛ̀sì/ is Igala /òhì/ 'answer').

Besides consonants, there are seven oral vowels: [i e ɛ a ɔ o u]; in addition, [ɪ ʊ] occur in Igala and some Yoruba dialects. Nasal vowels are also attested. As shown in Table 2, dialects of Yoruba have four to six nasal vowels, Itsɛkiri has three nasal vowels, but Igala has no nasal vowels.

Oral vowels exhibit co-occurrence restrictions. For instance, /ɛ/ and /ɔ/ do not co-occur with /e/ and /o/ in nouns: ɔ̀bɛ 'knife', ɛbɔ 'sacrifice', ekpo 'palm-oil', ɔ̀kpó 'pillar'. Consonant–vowel co-occurrence restrictions are also attested. For example, /r/, /ɣ/, /w/, and /h/ are nasalized when nasal vowels occur after them: ɪyɔ̃̀ 'pounded yam', irĩ̀ 'iron', ìwɔ̃̀ 'measurement', ah̃ɔ̃̀ 'tongue'.

There are three tones (pitches): High (´), Low (`), and Mid (unmarked). These tones are crucial for understanding the phonology and syntax (sentence structure) of YRB languages. For example, tone can distinguish words: bí 'give birth', bì 'vomit', bɪ 'ask', arò 'lamentation', aro 'cymbal', aró 'indigo', àró 'a granary'. Tone may also serve a grammatical function. For example, within a statement, a high tone must occur between the subject and the verb: Dajɔ̀ (a name); Dayɔ̀ ɔ́ ra òrombó 'Dayo bought (some) oranges'.

Tone is used in other contexts. For instance, it is used in whistled Yoruba, a language used to communicate over long distances on farms. As speakers talk and whistle simultaneously, the language is transformed: consonants are devoiced (produced without vocal cord vibration) or turned to [h] and all vowels are changed to [u] because whistling involves lip puckering, the articulatory gesture required for the production of [u]. However, all tones are retained without any alteration. The retention of tones enables speakers to understand the meaning of whistled language. For instance, a farmer just arriving on the farm may greet farmers in nearby farms using any of the whistled forms in (2) and (3). As shown below, although the consonants and vowels of these whistled forms are altered, they have the same meaning as the nonwhistled form in (1):

(1) Nonwhistled Yoruba: ara iwájú, ɛ kú iʃɛ́ o 'people who arrived before me, greetings at work'

(2) Whistled form with devoiced consonants: u̥rú u̥w̥ú̥ɟú u kú u̥ʃú u

(3) Whistled form with [h] replacement: uhú uhúhú, u hú uhú u

The Yoruba talking drum, a pressure drum, which accompanies singing during festivals and important ceremonies, also uses tone. This drum 'speaks' by reproducing the tones of vowels. For example, a

TABLE 1 Consonants

	Labial	Alveolar	Palatal	Velar	Labial–Velar	Glottal
Stop						
Voiceless		t		kʷ	k͡p	
Voiced	b	d	ɟ	gʷ	g͡b	
Fricative				.		h
Voiceless	f	s	ʃ			h
Voiced				ɣ		
Nasal	m	n				
Liquid		l r				
Glide			y		w	

TABLE 2 Nasal Vowels

Yoruba	Itsɛkiri	Igala
Iɟeʃa: ĩ ɪ̃ ɛ̃ ɔ̃ ʊ̃ ũ	ĩ ɔ̃ ũ	No nasal vowels
Ilaɟe: ĩ ɛ̃ ã ɔ̃ ũ		
Standard Yoruba: ĩ ɛ̃ ɔ̃ ũ		

notable citizen who has just returned from a trip may be greeted with these words using drums:

(4) Words: a ti ń retí ìrɛ; káábɔ̀, ʃé dáadaa lo dé 'we have been expecting you; welcome, did you arrive well?'

Talking drum: Mid Mid High Mid High Low Mid, High High Low High High Mid High Mid Mid High Because the talking drum produces only tones, its language could be ambiguous. For instance, the signature tune of Radio Nigeria Ibadan, given in (5), could be assigned the interpretations in (6):

(5) Talking drum: High High Low Low High High Mid High High Mid Low

(6) Possible interpretations

a. rédó nàìɟíríá la tí ń fɔhṹ
'we are broadcasting from Radio Nigeria'

b. tólúbàdɔ̃̀ bá kú ta ní ó ɟoyè
'if the king of Ibadan dies, who will succeed him?'

c. nǐnṹ kòkò dúdú la ti ń sebè,
'we cook soup in a black pot'

Syllables may be expressed as Consonant–Vowel (lɔ 'go'), Vowel (ilé 'house'), or N, a syllabic nasal (òntè 'stamp'). Syllables with consonant clusters and those ending in consonants are unattested. Hence, when English words with such unacceptable sequences are borrowed, they are restructured through consonant deletion and vowel insertion, as follows:

(7)

English	YRB
kettle	kɛ́tù
buckle	bɔ́kù
travel	tíráfù
brick	bíríkì

The Consonant–Vowel syllable plays an important role in Yoruba phonology. For instance, ɛnɔ̃̀, a Yoruba secret language game, disguises words by adding a Consonant–Vowel syllable to the end of every syllable. As shown in (8), the consonant of the added syllable is fixed—a/g/; however, the vowel and tone of this syllable vary, based on the vowel and tone of the base form. Another property of this game is that the final Consonant–Vowel tag is always preceded by a syllabic nasal:

(8) Nondisguised form: olú kò lɔsílé
'Olu did not go home' Olu NEG. go to house
Disguised form: ogolúgú kògò lɔgɔ sígíléŋ́gè
'Olu did not go home'

Word-initial high tones are also sensitive to syllable shapes. For example, whereas a word-initial Vowel or Consonant–Vowel syllable may bear a high tone in Igala, only Consonant–Vowel syllables may have a high tone in Yoruba and Its ɛkiri . Hence, vowel-initial words with high tones in Igala surface with mid tones in Yoruba:

(9)

Yoruba	Igala	Gloss
imú	ímɔ	nose
etí	étí	ear
fújɛ́	fújɛ́fújɛ́	to be light
tǐrǐ̃	fíílí	narrow, thin

Morphology (Word Structure)

Prefixation, reduplication, truncation, compounding, and desentencing are the principal word formation devices. Although inflection does not play a role in Yoruba and Igala, Itsɛkiri distinguishes between singular and plural nouns, as shown below:

(10)

Yoruba	Igala	Itsɛkiri	Gloss
obĩrĩ	ónobùlɛ	obirɛ̃̀	woman
obĩrĩ	ónobùlɛ	ebirɛ̃̀	women

Prefixation (the process of attaching an affix to a word to derive a new word) is a device used to derive nouns from existing verbs. For example, a vowel may be prefixed to a verb or verb phrase to form a noun, as follows:

(11) à−lɔ 'going' <à+lɔ 'prefix +go'

ɔ− 'singer' <ɔ+ 'prefix
kɔrĩ kɔrĩ +sing'

Reduplication (process of repetition) is used to form words denoting intensity; it is also used to form distributive nouns, profession-based names, and names denoting endearment:

(12) Base   Reduplicated Form

Intensity:

kíá	'quick'	kíákíá	'quickly'
kpú	'much'	kpúkpɔ́	'very much'
kpɔ́		kpúkpɔ́	

Distributive:

oʃù	'month'	oʃooʃú	'every month'
àg	'elder'	àgbà	'every elder'
bà		àgbà	

Profession agentive:

wolé	'examine house'	woléwolé	'sanitary inspector'
mɔ̃́	'build house'	mɔ̃́lé	'builder'
lé		mɔ̃́lé	

Diminutives:

| adé | 'a name' | ádéadè | 'diminutive form' |
| olú | 'a name' | ólúolù | 'diminutive form' |

Truncation (shortening or abbreviation) is the device adopted in creating shortened names, which are commonly used in informal contexts:

(13)

Full Name	Shortened Form
adéwɔlé	adé or wɔlé
olúkɔ́lá	olú or kɔ́lá
ɔlábɔ̀dé	ɔlá or bɔ̀dé
modúkpɛ́	modú or dúkpɛ́

Compounds (words consisting of two or more words) are formed from two or more nouns, as follows:

(14)

ìdí kɔ̀	'motor park'	<ì dí	'base/ bottom'	+ɔkɔ̀	'vehicle'
ɛrɔ̃̀ ko	'animal'	<ɛrɔ̃̀	'meat'	+oko	'farm'
ɔmɔ̃́ bĩrĩ̀	'girl'	<ɔmɔ̃́	'child'	+obĩrĩ̀	'woman'
ɔmɔ̃́ kas è	'toe'	<ɔmɔ̃́	'child'	+ì ka +ɛsɛ̀	'finger' 'leg'

A rather different but common morphological process in YRB and Benue-Congo languages is desentencing, a process used to form names. Desentencing involves turning a full sentence into a name, as follows:

(15)

Sentence	Name	Gloss
ɔlá wɔ ilé	ɔláwɔlé	honor enters the house
adé yí mi ká	adéyíŋká	crown surrounds me
akĩ̀ dé ilé	akĩ̀déilé	valor arrives in the house
olú dá mi sí	olúdámisí	the Lord kept me alive

Syntax (Sentence Structure)

Nouns, verbs, and prepositions are the major word categories. Some words, which function as adjectives and adverbs in English, are expressed as verbs. For example, the adjective 'red' and the adverb 'again' are used as verbs in the following Yoruba sentences:

(16)

| olú | pukpa | 'Olu is red: Olu is fair-skinned' |
| Olu | red | |

| olú | tũ̀ ń sáré | 'Olu is running again' |
| Olu | repeat is run | |

The basic word order is S(ubject–)V(erb–)O(bject). Within phrases, the head of the phrase occurs in the initial position. Hence, in a noun phrase, the noun occurs first, followed by adjectives and relative clauses. In verb and prepositional phrases, the verb and preposition appear first and their objects occur after them:

(17)

a.	Word order:	olú ra iwé Olu buy book	'Olu bought a book'
b.	Noun phrase:	ɔmɔ̃̀ dúdú child black	'black child'
c.	Verb phrase:	ra iwé buy book	'buy a book'
d.	Prepositional phrase:	sí oko to farm	'to the farm'

The basic SVO word order may be altered to form focus sentences, sentences in which a given word or phrase is fronted for emphatic purposes. In some Yoruba dialects, a focused word or phrase is usually followed by the focus marker *ni*; in other dialects, the focus marker appears at the end of the sentence; in Its Ekiri, only fronting applies, and the focus marker is not used. The following examples illustrate subject fronting, derived from a basic sentence such as (17a):

(18)

| Standard Yoruba Subject Focus: | olú ni Olu focus | ó ra iwé 3SG buy book | 'It is Olu who bought a book' |

(19)

Ukarɛ Yoruba:	olú Olu	ó ra iwéaĩ̃ 3SG	'It is Olu who bought a book'
Subject Focus:		buy book focus	
2	3		4

(20)

| Itsɛkiri Subject Focus: | olú òũ̀ é ra iwé Olu 3SG buy book | | 'It is Olu who bought a book' |

As shown in (18) through (20), a pronoun (3SG) replaces a fronted subject. In contrast, a fronted object does not produce the same effect:

(21) Standard
Yoruba
Object   ìwé ni olú rà   'It is a book
Focus: that Olu
                                  bought'
                                  book focus
                                  Olu buy

Serial-Verb-Construction (SVC) is another principal point of syntactic interest. SVCs have sequences of verbs, which share the same subject and object. They are different from coordinate clauses in that the verbs use one tense and aspect, and allow only one negative:

(22) a. olú   ra àgbàdo ɟɛ   'Olu will buy
     yóó   some corn and
                   eat them'
                   Olu FUT
                   buy corn eat

     b. olú   ra àgbàdo ɟɛ   'Olu did not
        kò   buy and eat
                                       some corn'
                   Olu NEG
                   buy corn eat

As shown below, it is ungrammatical to assign separate tense or negative markers to a serial verb construction:

(23) *olú yóó   'Olu will buy some
     ra àgbàdo yóó ɟɛ   corn and eat them'

     Olu FUT buy corn
     FUT eat

     *olú kò,   'Olu did not buy and eat
     ra àgbàdo   some corn'
     kò ɟɛ
     Olu NEG buy corn
     NEG eat

YRB languages typically divide pronouns into two sets, long and short pronouns:

(24)
Yoruba		Igala		Gloss
Long	Short	Long	Short	
èmi	mo, m	omi	ù, ǹ	I
ìwɔ	o	ùwE,	ɛ̀	You
òũ	ó	òũ	ì	He/she/it

Long pronouns are also called independent pronouns because they behave like nouns in terms of fronting and coordination, as shown by these examples:

(25) Basic   ìwɔ́ ra   'you bought
     Sentence: àgbádá   a flowing
                                  gown'

     Subject   ìwɔ ni o   'it is you
     Focus: ra àgbáda   who bought
                                  a flowing
                                  gown'

     Coordination: ìwɔ àti o   'you and
                   lú ra à   Olu bought
                   gbáda   a flowing
                                   gown'

Unlike long pronouns, short pronouns can neither be fronted nor coordinated with a noun.

Politeness

Finally, politeness is an essential aspect of Yoruba, Itsɛkiri, and Igala culture, and language is used to convey respect, especially for older people and those in authority. Politeness is expressed by the choice of pronouns in reference and address. For instance, the plural forms of the second and third person pronouns (èyĩ 'you pl.,' àwɔ̃ 'they') are considered respectful and appropriate whereas singular forms are considered impolite (ìwɔ 'you sg.,' òũ 'he/she').

References

Abraham, Roy C. 1958. Dictionary of Modern Yoruba. London: University of London Press.

Awobuluyi, Ọládele. 1978. Essentials of Yoruba grammar. Ibadan: Oxford University Press.

Bamgbose, Ayọ́. 1967. A short grammar of Yoruba. Ibadan: Heinemann Publishers.

Capo, Hounkpati B.C. 1989. Defoid. The Niger-Congo languages, ed. by J. Bendor-Samuel. Lanham: University Press of America.

Fresco, Edward. 1970. Topics in Yoruba dialect phonology. Studies in African linguistics, Supplement to Vol. 1. Los Angeles: University of California Press.

Rowlands, Evan C. 1969. Teach yourself Yoruba. London: English Universities Press.

Schleicher, Yetunde. 1993. Jẹ ka sọ Yoruba (let us communicate in Yoruba). New Haven: Yale University Press.

Williamson, Kay, and Roger Blench, Niger-Congo. 2000. African languages: an introduction, ed. by Bernd Heine, and Derek Nurse. Cambridge: Cambridge University Press.

ỌLANIKẸ ỌLA ORIE

Z

Zipf, George Kingsley

A philologist who studied and taught German and linguistics at Harvard University for most of his academic career, George Kingsley Zipf was one of the first scholars to use statistics in the study of language. Although he is perhaps best known today for the statistical formula bearing his name, Zipf did not actually create the formula; it was used earlier in the writings of such scholars as J.B. Estoup and Vilfredo Pareto. However, Zipf served as its most ardent proponent, applying the formula to language and a variety of other social phenomena in his argument for a human ecology.

Zipf began his work on the relationship between frequency and language as a graduate student at Harvard, where he successfully defended a dissertation on the effect of relative frequency on phonetic change. In this work, Zipf argued for a Principle of Relative Frequency, which said that

> the accent or degree of conspicuousness of any word, syllable, or sound is inversely proportionate to the relative frequency of that word, syllable, or sound, among its fellow words, syllables, or sounds in the stream of spoken language. As any element's usage becomes more frequent, its form tends to become less accented, or more easily pronounceable, and vice versa. (1932:1)

Using the Chinese, French, and English languages, as well as a number of writings in Latin, Zipf addressed the correlation between the size of a word and its frequency in a language, finding that the size of words in the language is inversely proportional to its frequency in the language, so that short words will occur very often in the language and long words will occur infrequently. He also argued that the use of abbreviations among speakers followed this same trend. Zipf's explanation for the correlation between the two was that 'High frequency is the cause of small magnitude' (1935:29) and he attributed this correlation to a Principle of Least Effort, which he said 'means … that a person … will strive to solve his problems in such a way as to minimize the total work that he must expend in solving both his immediate problems and his probable future problems…' (1949:1).

To explain, Zipf used the analogy of an artisan's workbench. According to this analogy, the tools of an artisan that are used more often will be placed nearer the artisan's chair so that they may be retrieved with the least possible effort on the artisan's part, while the tools that are rarely used will be placed farther away from the artisan.

Following the same line of reasoning, Zipf also worked on the relationship between rank and frequency of words in a language. Zipf found that the frequency of a work in a language is *inversely proportional* to the rank of that word in the language, so that if a word ranks first in the language, it will be used far more frequently than those words ranked lower in the language. Conversely, a word that ranks 100th in the language will be used far more rarely than words that rank closer to one in the language. In the form of an equation, this takes the form of Rank times Frequency equals Constant.

Zipf later applied his formula to the populations and ranks of city with some success. In the introduction to his National Unity and Disunity (1941), Zipf explains that if readers were to look at the US Census in any given year they would find the city with the greatest population to rank first, the second-ranking city would have half the

population of the first, the third-ranking city would have one third the population, and so on. Zipf added that the size and rank of incomes would also follow this simple mathematical formula. Zipf inferred from this that communities—cities as well as nations—worked by the same principles as other organisms, so that we should think of a population as a 'natural biosocial entity'.

From the beginning, Zipf's work has been the subject of much criticism. Perhaps greatest among these concerns is Zipf's attempt to explain his findings with the Principle of Least Effort, due to the significant problem of identifying articulatory features that are universally 'difficult' or 'easy'. Another criticism that many people have made—and one that is supported by Zipf's evidence of the formula working in fields besides linguistics—is that Zipf's Law does not succeed so much at describing how language works as it does at describing how mathematics works. Finally, although Zipf's Law typically works well with respect to objects that are ranked in the middle of lists, it is less dependable with objects ranked very low or very high, a problem that has led some, such as mathematician Benoit Mandelbrot in his work on fractal geometry (1982), to make some modifications when using the formula.

Although there are problems with Zipf's Law, and certainly with the inferences that Zipf made from his formulations, Zipf should be recognized as one of the pioneers of the use of statistical analyses in linguistics in general and the current subfield of corpus linguistics in particular. Zipf's Law is also commonly used in mathematics, physics, library science, and memetic science, which is the quantitative analysis of cultural transfer. Zipf died on September 25, 1950 in Newton, Massachusetts, after a long illness.

Biography

Born in Freeport, Illinois, on January 7, 1902, George Kingsley Zipf received his B.A. (1924) and his Ph.D. (1930) in Comparative Philology from Harvard College, completing a dissertation on relative frequency of use as a cause of phonetic change in language evolution. He was Instructor of German at Harvard until 1936, Assistant Professor of German until 1939, and University Lecturer until 1950. He spent the year after his graduation from Harvard studying at the University of Bonn and the University of Berlin. He was Member of Linguistic Society of America (LSA), 1931, and served on the Committee on Quantitative Linguistics, created by the Sixth International Congress of Linguists, in 1948. He died in Newton, Massachusetts, on September 25, 1950.

References

Mandlebrot, Benoit, B. 1953. An information theory of the statistical structure of language. Communication theory, ed. by W. Jackson, 486–502. London: Butterworths.

——— 1982. The fractal geometry of nature. San Francisco: W.H. Freeman and Co.

Miller, George. 1965. Introduction. The psycho-biology of languages: an introduction to dynamic philology, ed. by George Zipf, v–x. Boston: MIT.

Rapoport, Anatol. 1982. Zipf's law revisited. Studies on Zipf's law, ed. by H. Guiter and M.V. Arapov. Bochum: Studienverelag Brockmeyer.

Walsh, J.L. 1949. Another contribution to the rapidly growing literature of mathematics and human behavior. Scientific American, August. 181(2). 56–8.

Zipf, George Kingsley. 1929. Relative frequency as a determinant of phonetic change. Harvard Studies in Classical Philology, XL.

———. 1932. Selected studies of the principle of relative frequencies in language. Havard University Press.

———. 1935. The psycho-biology of language: an introduction to dynamic philology. Boston: Houghton Mifflin.

———. 1941. National unity and disunity: the nation as a biosocial organism. Bloomington, IN: Principia Press.

———. 1949. Human behaviour and the principle of least effort: an introduction to human ecology. Cambridge: Addison-Wesley.

LAMONT ANTIEAU

Zulu and Southern Bantu Languages

The Sintu (Bantu) languages are the most widespread and best-known subgroup of Niger-Congo (the form Sintu reflects the Nguni prefix isi-, denoting language and culture). According to Greenberg's (1963) classification, the Niger-Congo is a subfamily of the Niger-Kordofanian family, one of the four language families in Africa. The Sintu languages cover most of the vast southern third of Africa, from the equatorial rain forest of Gabon and southern Cameroon to the highlands of Kenya and from the Ubangi River in the Central African Republic nearly to the Cape of Good Hope in South Africa. Some of the well-known Sintu languages include Zulu, Kongo, Lingala, Shona, Bemba, Ganda, and Gikuyu.

Despite their vast extent today, Sintu languages belong to just one subgroup of the Benue-Congo branch of Niger-Congo (Williamson and Blench 2000). Sintu languages spread eastward and southward from West Africa (modern-day eastern Nigeria and Cameroon) in the early centuries of the first millennium AD. They only spread over the regions they now occupy during the last 4,000 years. The ancestral Sintu language, proto-Bantu, was spoken somewhere probably in eastern Nigeria, West Africa. Then, for reasons unknown, early Sintu communities began expanding into new territories, first through the equatorial rain forest belt and along its margins, and then, between about 500 BC and AD 300, eastward and southward into eastern and southern Africa. It remains unknown what languages preceded the Sintu in the equatorial forest and the adjoining savannas, but in southern Africa the earlier languages often belonged to the Khoisan family.

As Sintu languages spread throughout Africa, their features remained remarkably stable over great distances and long periods of time. Their main features are the systems of noun classes and agglutinative verb morphology. Guthrie (1948) divided Bantu (Sintu) languages into six areas: northwest Central Africa, west and southwest Central Africa, east central, northeast central, southeast, and south. Some of the most important representative members of the southeast area include Shona, Tsonga, Ronga, Makua, and Yao; those of the south area include Swazi, Tswana, and Zulu-Xhosa. In Voegelin and Voegelin's (1997) similar classification of the world's languages, 'Bantu proper' is divided into seven major areas: central eastern, central western, Kari, northeastern, northwestern, southeastern, and southwestern areas. Languages of the southeastern area are divided into two groups: Makua, Matumbi, and Yao, with speakers in Tanzania, Mozambique, and Malawi; and Chopi, Nguni, Shona, Sotho-Tswana, Tswa-Ronga, and Venda, with speakers in Zimbabwe, Botswana, South Africa, Swaziland, and Lesotho. The Nguni include isiSwati, isiXhosa, Gaika, Gcaleka, Mpondo, Mpondomse, Ndlambe, and *isiZulu.* The southwestern area contains two subdivisions: Chokwe-Luchazi, Lozi, Luyana, Subiya, with speakers in Botswana, Zambia, Angola, and Zaire; and Herero, Ndonga, Umbundu, and Yeye reside in Angola, Namibia, and Botswana.

The languages that Guthrie (1948) recognized as Bantu are now referred to as 'Narrow Bantu' (Williamson and Blench 2000). All major modern attempts to clarify Narrow Bantu have not produced an agreed overall scheme (Nurse 1996). The most widespread agreement is that there is a northwest Bantu (Sintu); some authors see a division between east and west Bantu (Sintu). The southern Sintu languages include Xhosa, Zulu, Sotho, Tsonga, and Tswana. The

high degree of structural unity among most Sintu languages, the wide area of contact among them, and significant mutual influence among Sintu languages make classification difficult.

The most typical feature of Sintu morphology is the class prefix system (Campbell 1991). Proto-Bantu had 19 noun classes, while Sotho, for example, has only seven. Sintu languages have long been appreciated by scholars for their distinctive morphology, being highly agglutinative, and allowing great structural complexity to nominal and even more so to verbal forms. There are very few root adjectives in Sintu; some examples are –*kulu* (large in *isiZulu*) and –*bi* (bad in *isiZulu*). The conjunctive pronouns, subjective and objective, are remarkably homogeneous over most of the Sintu area. Most primary verb roots are disyllables, and derived stems are formed by suffixation. For example, bonana means 'to see each other' in *isiZulu*. Some Sintu languages have a neutral passive of state ub –V*ka(la);* for example, in *isiZulu inkanyezi ya.bona.ka.la* ('the star was visible'). Moods are generally marked by suffix. Most Sintu languages have seven moods: infinitive, indicative, imperative, subjunctive, perfect, continuative, and relative. In terms of syntax, the typical Sintu verbal complex consists of prefix (subject concord marker)–tense marker–object marker–stem–modal/voice marker.

Current thinking is that the Nguni were one of three large African migrant groups whose tradition of horticulture and cattle breeding combine the major cultural attributes of West, Central and North East Africa, from where they moved along separate routes to southern Africa. The Nguni followed an inland course via the headwaters of the Zambesi, where contact with San hunters produced the 'click' sounds that characterize their languages today. They moved southward to the most northerly bend of the Limpopo River, which marks the boundary between present-day South Africa and Zimbabwe, sometime between the middle of the fifteenth and the third century AD. The Nguni supposedly split into separate migrations, moving in different stages into what is now called KwaZulu-Natal and the Eastern Province (formerly known as the Transkei). They fell broadly into four regional divisions: the Mhlubi, Mtetwa, Ndwandwe, and Ngwane.

Some of those who settled in northern KwaZulu-Natal doubled back into what is now Swaziland, while those who first entered the Transkei were the forebears of the Pondo. The last to leave the Limpopo settled for a while in what is now the southeastern region of the Mpumalanga province; and then they moved on in easy stages into central KwaZulu-Natal. Finding the northeast and northwest already occupied, two smaller groups moved on. One of these, finding the coastal regions of the south settled by the Pondo, kept to the

inland high ground, to become the Xhosa. The other of the two smaller groups found a home as the coastal neighbors of the Pondo to become the Thembu of today. Thus, the Nguni divided into two large segments, North and South. The Xhosa, Pondo, and Thembu of the Eastern Cape (formerly Transkei) and areas further south are major representatives of the South Nguni, while the Zulu, the Swazi of Swaziland, and the Ndbele (in the present provinces of Gauteng and Mpumalanga) are of the Northern Nguni.

The final Nguni migration populated the heart of KwaZulu-Natal where the small and unimportant Zulu clan was later to succeed the Ndwandwe and Mthetwa empires, respectively, in the northwest and northeast. Under their famous chief, Shaka kaSenzangakhona (c. 1787–1828), the first kind of this unified group of Nguni clans and chiefdoms, they became the rulers of the northern half of KwaZulu-Natal from the Tugela River in the south to the border of Mozambique in the north. A long line of descendants link these historic figures with the current royal house headed by King Goodwill Zwelethini.

Prior to the Zulu king Shaka, Zulu referred strictly to a relatively small clan of people living in the middle reaches of the White Mfolozi River. When Shaka founded the Zulu kingdom in the early nineteenth century, he extended the use of Zulu to the many clans he incorporated into the Zulu state. European anthropologists and linguists, who saw the Zulu language as central in defining identity, categorized people of Natal and Zululand as Zulus (Harries 1993).

The Nguni languages are closely related, and now spoken by most people in South Africa. On purely linguistic grounds, there are two main languages in southern Africa that have severally mutually intelligible dialects: Nguni (with varieties like Ndebele, Swazi, Xhosa, Zulu) and Sotho (northern Sotho, southern Sotho, Tswana) (Wolff 2000). In particular, *isiZulu* is closely related to Swazi and Xhosa. However, members of these speech communities, for several reasons, consider themselves to have separate languages, not just dialects, and unique historical, ethnic, and social backgrounds (Campbell et al. 1995). Language is closely linked with ethnic identity in South Africa.

Outside Khoisan, clicks are known to occur only in South African Sintu languages such as Zulu, southern Sotho, Yeyi, and Xhosa, and in the Cushitic language Dahalo. A click is a multiply articulated sound produced by forming one closure in the front of the mouth with the lips of the tongue in front and another in the back of the mouth with the tongue dorsum. The clicks are a result of borrowing from Khoisan by way of long-enduring and intensive contacts, rather than a genuine part of the sound inventories in Sintu languages themselves. For example, the Zulu click sounds consist of a

dental, palatal, and lateral click, each containing four phonemes—a surd, its aspirate, its voiced allophone, and its nasal (Campbell 1991:163).

IsiZulu is now the best-known language in South Africa, besides English. In 1995, there were 9,140,000 Zulu people in southern Africa. They consisted of 8,778,000 Zulu people in South Africa (about 18.8% of the population), 37,480 in Malawi, 76,000 in Swaziland, and 248,000 in Lesotho (Grimes 1996). Now about 22.4% of the South African population use it as their home language, while others speak it as a second language.

IsiZulu was first studied by foreign linguists in the nineteenth century. Natal's first Anglican bishop, John William Colenso, and his assistant, William Ngidi, translated and studied the Zulu language (*isiZulu*) in the 1850s and early 1860s (Guy 1997). *IsiZulu* is a tonal language, closely related to Swazi and Xhosa. Its two dialects are Lala and Qwabe. The standardization of *IsiZulu* is now controlled from Kwa-Zulu.

There are also several slang languages that are variations of *isiZulu*. For example, Tsotsitaal, Iscamtho, and Fanagolo are slang languages that are based on Afrikaans, English, and Zulu. Fanagolo is a second language only, with no mother tongue speakers, widely used in towns and gold, diamond, coal, and copper mining areas, while Tsotsitaal and Iscamtho are spoken in black urban townships in South Africa (Slabbert and Myers Scotton 1997).

References

Bendor-Samuel, John (ed.) 1989. The Niger-Congo languages: a classification and description of Africa's largest language family. New York: University Press of America.

Bryant, A. T. 1929. Olden times in Zululand and Natal. London: Longmans; reprint, Cape Town: C. Struik, 1965.

Campbell, C., G. Mare, and C. Walker. 1995. Evidence for an ethnic-identity in the life-histories of Zulu-speaking Durban township residents. Journal of Southern African Studies 21(2). 287–301.

Campbell, George L. 1991. Compendium of the world's languages. London: Routledge.

De Kadt, E. 1998. The concept of face and its applicability to the Zulu language. Journal of Pragmatics 29(2). 173–91.

Greenberg, Joseph H. (ed.) 1963. Universals of language. Cambridge, MA: MIT Press.

Grimes, Barbara F. (ed.) 1996. Ethnologue. Summer Institute of Linguistics.

Guthrie, M. 1948. Classification of the Bantu languages. Oxford: Oxford University Press.

Guy, J. 1997. Class, imperialism, and literary criticism: William Ngidi, John Colenso and Matthew Arnold. Journal of Southern African Studies 23(2). 219–41.

Harries, Patrick. 1993. Imagery, symbolism and tradition in a South African Bantustan: Mangosuthu Buthelezi, Inkatha, and Zulu history. History and Theory 32(4). 105–26.

Nurse, Derek. 1996. 'Historical' classifications of the Bantu languages. The growth of farming communities in Africa

from the Equator southwards, ed. by J.E.G. Sutton. Nairobi: British Institute in Eastern Africa.

Slabbert, Sarah, and Carol Myers-Scotton. 1997. The structure of Tsotsitaal and Iscamtho: code switching and in-group identity in South African Townships. Linguistics: An Interdisciplinary Journal of the Language Sciences 35(2). 317–43.

Voegelin, C.F., and F.M. Voegelin. 1997. Classification and index of the world's languages. New York: Elsevier.

Williamson, Kay, and Roger Blench. 2000. Niger-Congo. African languages: an introduction, ed. by Bernd Heine and Derek Nurse. Cambridge: Cambridge University Press.

Wolff, H. Ekkehard. 2000. Language and society. African languages: an introduction, ed. by Bernd Heine and Derek Nurse. Cambridge: Cambridge University Press.

CAMILLA M. COCKERTON

See also **Niger-Congo**

Zuni

The Zuni language spoken in the Pueblo region of the southwestern United States is a linguistic isolate as yet unproven to be related to any other known language. According to suggestive archeological records, Zuni land claims, and references to sites in oral tradition, the territory in which Zuni was originally spoken extended over much of what is now the four-corner region of New Mexico, Arizona, Utah, and Colorado. Today, Zuni is spoken at Zuni Pueblo and its associated villages and farming communities (Nutria, Ojo Caliente, Blackrock, Pescado) in northwest New Mexico, approximately 150 miles west of Albuquerque.

The Pueblo region is recognized as an important area of shared culture and history encompassing four other language families native to North America in addition to Zuni (Keresan, Kiowa-Tanoan, Uto-Aztecan, and more recently Athapaskan). Despite several attempts, genetic linguistic affiliation has not successfully been demonstrated between Zuni and any other language. One such attempt proposed a link to the Penutian languages of California. Subsequent investigation of the evidence presented has discredited this proposal, although the mistaken categorization of Zuni as Penutian has persisted.

Zuni differs from its closest geographical linguistic neighbors in many typological respects, for example, in lacking much of the complex verbal morphology typical of languages of this area as well as their associated free word order. Zuni has a basic word order of S(ubject)–O(bject)–V(erb), but special conditions on the position of pronouns or on words containing contrastive focus frequently result in OSV order. Evidence of loanwords in Zuni from Keresan (e.g. Zuni *k'yaššita*, Acoma Keresan *sk'a'šу* 'fish'; Zuni *pu:la*, Acoma Keresan *pú:r'ai'ką* 'butterfly'; Zuni *'uwakya*, 'great-grandson' Acoma Keresan *'úwa'ką* 'baby') suggests a history of contact between Zuni and the Pueblos further to the east along the Rio Grande. The effect of this con-

tact on Zuni grammatical structure, however, is less apparent. Influence on grammatical structure as a result of contact with Uto-Aztecan speakers to the west is more easily detected. For instance, the system of clause linkage indicating continued vs. changed events along with a continued vs. change in sentence subject is constructed in similar fashion to that of Hopi, currently spoken in Arizona several hundred miles away.

While a great deal of archeological and anthropological work has been done at Zuni over the past century, the Zuni language has received less attention than other languages of Native North America, possibly due to its isolate status. Early documentation includes word lists collected by Catholic missionaries. Transcriptions of letters, stories and prayers from the nineteenth century have also been found in archives. Brief grammatical sketches were published in the 1930s and 1960s, as well as a short lexicon of words and grammatical endings. An orthography was constructed in the 1970s that is currently in use at the Pueblo for citation of Zuni words and names; however, Zuni in daily life largely remains an unwritten language. The orthography does not strictly follow standard linguistic practices for transcription; for example, it transcribes the voiceless stops [p, t] as 'b' and 'd', respectively, and it reduces some geminate (doubled) consonants to single consonants.

Among the interesting phonetic properties of Zuni are the glottalized series of consonants and the correlation between word stress and vowel tensing. (Word stress falls on the initial syllable, and the further a vowel occurs from this stress the laxer its pronunciation.) The phonological structure allows only a single consonant at the beginning of a word, only two consonants contiguous in the middle of a word, and only a subset of consonants at the end of a word. The language contains both long vowels and consonants (geminates), and adjacent consonants frequently assimilate to result in a

geminate, both historically as well as in active pronunciation. Similar to English, Zuni has a nominative/accusative case system and inflects only pronouns for case. Unlike pronominal elements of other languages in the geographical area, Zuni pronouns are obligatory independent forms and not inflectional morphology on the verb. Interestingly, while Zuni does not allow for pronouns to be dropped, it has no third person pronouns. Consequently, a verb appearing on its own receives third person anaphoric interpretation for its subject (and object if it takes one), cf. *'awakya,* find-past, 'he/she found him/her/it'. Zuni noun stems are marked with different inflectional endings in the singular and in the plural (*te'le* 'pot' vs. *tewe* 'pots'), and nouns have endings that classify the noun based partly on semantic classes and partly on the number of syllables of the stem (*no'le* 'bean' vs. *no'linne* 'nose').

Syntactic properties of broader typological interest include noun incorporation, which takes the special phonological form of a noun reduced to its initial consonant and vowel and attaches this reduced form as an affix to the verb. Furthermore, Zuni makes abundant use of sentence initial particles to indicate a range of meaning from aspect to modality. Zuni lacks the ability to form finite subordinate clause structures of the kind selected by verbs of attitude evaluation, such as the equivalent of 'I think that I will go' and instead uses this finely detailed system of sentence initial particles to convey attitude.

The population of Zuni Pueblo is currently around 10,000+, and it is estimated that 70–80% of speakers use Zuni as a first language. Many of the oldest generation of speakers are bilingual in Spanish, while English has been increasing as second and first language among younger generations.

References

Bunzel, Ruth. 1932a. Zuni origin myths. Report of the Bureau of American Ethnology. Smithsonian Institution, Washington, DC. Reprinted in Zuni Ceremonialism. 1992. Albuquerque: University of New Mexico Press.

———. 1932b. Zuni ritual poetry. Report of the Bureau of American Ethnology. Smithsonian Institution, Washington, DC. Reprinted in Zuni Ceremonialism. 1992. Albuquerque: University of New Mexico Press.

———. 1933. Zuni texts. Publications of the American Ethnological Society, Vol. 15. American Ethnological Society. Washington D.C.

———. 1933. Zuni. Handbook of American Indian languages, Part 3, ed. by F. Boas, 389–515. New York: J.J. Augustin.

Newman, Stanley. 1958. Zuni dictionary. International Journal of American Linguistics 24(1) (pt. 2).

———. 1965. Zuni grammar. University of New Mexico Publications in Anthropology, No. 14. Albuquerque: University of New Mexico Press.

LYNN NICHOLS

Index

Entries containing suffix '*f*' and '*t*' along with page numbers indicate figure and table, respectively.

Index

Index

Index

Index

Index

Index

Index

Index

Index

Index

Index

Index

Index

Index

Index