Ten Lectures on Field Semantics and Semantic Typology

Jürgen Bohnemeyer - 978-90-04-36262-8
Downloaded from Brill.com10/03/2021 01:49:08AM
via Massachusetts Institute of Technology

Distinguished Lectures in Cognitive Linguistics

Edited by
Fuyin (Thomas) Li (*Beihang University, Beijing*)

Guest Editor
Yan Ding (*Beijing Jiaotong University*)

Editorial Assistants
Jing Du (*University of Chinese Academy of Sciences*), Na Liu and Cuiying Zhang (*doctoral students at Beihang University*)

Editorial Board
Jürgen Bohnemeyer (*State University of New York at Buffalo, USA*) – Alan Cienki (*Vrije Universiteit (VU), Amsterdam, Netherlands and Moscow State Linguistic University, Russia*) – William Croft (*University of New Mexico, USA*) – Ewa Dąbrowska (*Northumbria University, UK*) – Gilles Fauconnier (*University of California at San Diego, USA*) – Dirk Geeraerts (*University of Leuven, Belgium*) – Nikolas Gisborne (*The University of Edinburgh, UK*) – Cliff Goddard (*Griffith University, Australia*) – Stefan Th. Gries (*University of California, Santa Barbara, USA*) – Laura A. Janda (*University of Tromsø, Norway*) – Zoltán Kövecses (*Eötvös Loránd University, Hungary*) – George Lakoff (*University of California, Berkeley, USA*) – Ronald W. Langacker (*University of California, San Diego, USA*) – Chris Sinha (*Hunan University, China*) – Leonard Talmy (*State University of New York at Buffalo, USA*) – John R. Taylor (*University of Otago, New Zealand*) – Mark Turner (*Case Western Reserve University, USA*) – Sherman Wilcox (*University of New Mexico, USA*) – Phillip Wolff (*Emory University, USA*) – Jeffrey M. Zacks (*Washington University, Saint Louis, USA*)

Distinguished Lectures in Cognitive Linguistics publishes the keynote lectures series given by prominent international scholars at the China International Forum on Cognitive Linguistics since 2004. Each volume contains the transcripts of 10 lectures under one theme given by an acknowledged expert on a subject and readers have access to the audio recordings of the lectures through links in the e-book and QR codes in the printed volume. This series provides a unique course on the broad subject of Cognitive Linguistics. Speakers include George Lakoff, Ronald Langacker, Leonard Talmy, Laura Janda, Dirk Geeraerts, Ewa Dąbrowska and many others.

The titles published in this series are listed at *brill.com/dlcl*

Ten Lectures on Field Semantics and Semantic Typology

By

Jürgen Bohnemeyer

BRILL

LEIDEN | BOSTON

Brill has made all reasonable efforts to trace all rights holders to any copyrighted material used in this work. In cases where these efforts have not been successful the publisher welcomes communications from copyright holders, so that the appropriate acknowledgements can be made in future editions, and to settle other permission matters.

The Library of Congress Cataloging-in-Publication Data is available online at https://catalog.loc.gov
LC record available at https://lccn.loc.gov/2021031954

Typeface for the Latin, Greek, and Cyrillic scripts: "Brill". See and download: brill.com/brill-typeface.

ISSN 2468-4872
ISBN 978-90-04-36261-1 (hardback)
ISBN 978-90-04-36262-8 (e-book)

Copyright 2021 by Jürgen Bohnemeyer. Published by Koninklijke Brill NV, Leiden, The Netherlands.
Koninklijke Brill NV incorporates the imprints Brill, Brill Nijhoff, Brill Hotei, Brill Schöningh,
Brill Fink, Brill mentis, Vandenhoeck & Ruprecht, Böhlau Verlag and V&R Unipress.
Koninklijke Brill NV reserves the right to protect this publication against unauthorized use. Requests for re-use and/or translations must be addressed to Koninklijke Brill NV via brill.com or copyright.com.

This book is printed on acid-free paper and produced in a sustainable manner.

Contents

Note on Supplementary Material VII
Preface VIII
About the Author X

1 **Setting the Stage: Meaning, Cognition, Culture, and Crosslinguistic Variation** 1

2 **Field Semantics: Studying Meaning without Native Speaker Intuitions** 20

3 **Data Gathering in Linguistics: a Practical Epistemology of Elicitation Techniques** 37

4 **Sources of Evidence: Semantic and Pragmatic Diagnostics** 59

5 **Ethnosemantics and Cognitive Anthropology: a Short History** 79

6 **Semantic Typology: the Crosslinguistic Study of Semantic Categorization** 99

7 **Framing Whorf: Reference Frames in Language, Culture, and Cognition** 116

8 **Doing the Math: Quantitative Methods in Semantic Typology** 137

9 **Event Description: Variation at the Syntax-Semantics Interface** 157

10 **The Language-Specificity of Conceptual Structure: Taking Stock** 181

About the Series Editor 199
Websites for Cognitive Linguistics and CIFCL Speakers 200

Note on Supplementary Material

All original audio-recordings and other supplementary material, such as hand-outs and PowerPoint presentations for the lecture series, have been made available online and are referenced via unique DOI numbers on the website www.figshare.com. They may be accessed via a QR code for the print version of this book. In the e-book both the QR code and dynamic links will be available which can be accessed by a mouse-click.

The material can be accessed on figshare.com through a PC internet browser or via mobile devices such as a smartphone or tablet. To listen to the audio recording on hand-held devices, the QR code that appears at the beginning of each chapter should be scanned with a smart phone or tablet. A QR reader/scanner and audio player should be installed on these devices. Alternatively, for the e-book version, one can simply click on the QR code provided to be redirected to the appropriate website.

This book has been made with the intent that the book and the audio are both available and usable as separate entities. Both are complemented by the availability of the actual files of the presentations and material provided as hand-outs at the time these lectures were given. All rights and permission remain with the authors of the respective works, the audio-recording and supplementary material are made available in Open Access via a CC-BY-NC license and are reproduced with kind permission from the authors. The recordings are courtesy of the China International Forum on Cognitive Linguistics (http://cifcl.buaa.edu.cn/), funded by the Beihang University Grant for International Outstanding Scholars.

The complete collection of lectures by Jürgen Bohnemeyer can be accessed by scanning this QR code and the following dynamic link: https://doi.org/10.6084/m9.figshare.c.4792911

© JÜRGEN BOHNEMEYER, 2021 | DOI:10.1163/9789004362628_001

Preface

The present text, entitled *Ten Lectures on Field Semantics and Semantic Typology* by Jürgen Bohnemeyer is a transcribed version of the lectures given by Professor Jürgen Bohnemeyer, in June 2012 as the forum speaker for *the 10th China International Forum on Cognitive Linguistics*.

The China International Forum on Cognitive Linguistics (http://cifcl.buaa.edu.cn/) provides a forum for eminent international scholars to talk to Chinese audiences. It is a continuing program organized by some prestigious universities in Beijing. The following is a list of organizers for CIFCL10.

Host:
Li Fuyin (Thomas): PhD/Professor, Beihang University

Co-organizers:
Liu Shisheng: PhD/Professor, Tsinghua University
Gao Yihong: PhD/Professor, Peking University
Shi Baohui: PhD/Professor, Beijing Forestry University
Lan Chun: PhD/Professor, Beijing Foreign Studies University
Wang Lifei: PhD/Professor, University of International Business and Economics

Professor Jürgen Bohnemeyer's lecture series was mainly supported by *the Beihang Grant for International Outstanding Scientists* for 2012 (Project number: Z1267, Project organizer: Thomas Fuyin Li).

The text is published as one of the *Eminent Linguists Lecture Series*. The transcription of the audio, proofreading the text, and publication of the work in its present book form, has involved many people's strenuous inputs. The initial drafts were completed by the following postgraduate students, in the order from lecture 1 to 10: Jin Hui, Hu Xiaofang, He Fengfeng, He Junxiu, Li Xing, Zhou Weilu, Bo Shaoying, Du Jun, Fu Hongxing, and Li Heng. Li Heng had revisions for the whole book. Then we editors did the word-by-word and line-by-line revisions. To improve the readability of the text, we have deleted the false starts, repetitions, fillers like *now, so, you know, OK, and so on, again, of course, if you like, sort of*, etc. Occasionally, the written version needs an additional word to be clear, a word that was not actually spoken in the lecture. We have added such words within single brackets […]. To make the written version readable, even without watching the film, we've added a few "stage directions", in italics also within single brackets: […]. The stage direction describes what the speaker was doing, such as pointing at a slide, showing an object, etc. The

speaker, professor Jürgen Bohnemeyer, and his Research Assistant, Randi E. Moore, did the final revisions. The published version is the final version approved by the speaker.

The publication of this book is sponsored by the Humanities and Social Sciences Research Program Funds of the Chinese Ministry of Education (Number: 09YJA740010).

Thomas Fuyin Li
Beihang University (BUAA)
thomasli@buaa.edu.cn

Yan Ding
Beijing Jiaotong University (BJTU)
yanding@bjtu.edu.cn

About the Author

Jürgen Bohnemeyer, born in 1965, is Professor of Linguistics at the State University of New York at Buffalo, and previously worked at Bielefeld University, Germany, Tilburg University, the Netherlands, and the Max Planck Institute for Psycholinguistics, the Netherlands. He received his PhD from Tilburg University in 1998. The title of his dissertation is *Time Relations in Discourse: Evidence from a Comparative Approach to Yukatek Maya* (Bohnemeyer 1998). In 2003, he began teaching at Buffalo and obtained tenure in 2008. Bohnemeyer specializes in semantic typology, the cross-linguistic study of semantic categorization. This emerging field is at the intersection of semantics, cultural anthropology and cognitive psychology. Bohnemeyer's work focuses on the semantic typology of space, time, and event representation.

Some representative works include:

Bohnemeyer, Jürgen, 2002 *The Grammar of Time Reference in Yukatek Maya*. Munich: LINCOM.

Bohnemeyer, Jürgen, and Mary D. Swift, 2004 Event realization and default aspect. *Linguistics and Philosophy* 27(3): 263–296.

Bohnemeyer, Jürgen, 2007 Morpholexical Transparency and the argument structure of verbs of cutting and breaking. *Cognitive Linguistics* 18(2): 153–177.

Bohnemeyer, Jürgen, N. J. Enfield, James Essegbey, Iraide Ibarretxe-Antuñano, Sotaro Kita, Friederike Lüpke, and Felix K. Ameka. 2007 Principles of event segmentation in language: The case of motion events. *Language* 83(3): 495–532.

Bohnemeyer, Jürgen 2009 Temporal anaphora in a tenseless language. In Wolfgang Klein and Ping Li (eds.), *The expression of time in language*. Berlin: Mouton de Gruyter. 83–128.

Bohnemeyer, Jürgen, 2014 Aspect vs. relative tense: The case reopened. *Natural Language and Linguistic Theory* 32(3): 917–954.

Bohnemeyer, Jürgen, Katharine Donelson, Randi E. Moore, Elena Benedicto, Alyson Eggleston, Carolyn O'Meara, Gabriela Pérez Báez, Alejandra Capistrán Garza, Nestor Hernández Green, María de Jesús S. Hernández Gómez, Samuel Herrera Castro, Enrique Palancar, Gilles Polian, and Rodrigo Romero, 2015 The contact diffusion of linguistic practices: Reference frames in Mesoamerica. *Language Dynamics and Change* 5(2): 169–201.

Bohnemeyer, Jürgen and Robert D. Van Valin, Jr., 2017 The Macro-Event Property and the Layered Structure of the Clause. *Studies in Language* 41(1): 142–197.

Bohnemeyer, Jürgen, 2020 Linguistic relativity: From Whorf to now. In Gutzmann, D., L. Matthewson, C. Meier, H. Rullmann, & T. E. Zimmermann (eds.), *The Blackwell companion to semantics*. London: Blackwell.

Bohnemeyer, Jürgen, In press Elicitation and documentation of tense and aspect. *Language Documentation and Conservation* (2022).

LECTURE 1

Setting the Stage: Meaning, Cognition, Culture, and Crosslinguistic Variation

Thanks for having me. It's a great honor. It's in fact a great adventure for me. It's my first trip to China, my second trip ever to Asia, and the first one was to a very different part of Asia, to Turkey. In fact, I'd like to take you all on an adventure today and over the following four days. I'd like to take you on an adventure exploring the human mind. I'll be our tour guide, but our road map and our means of transport and even our means of exploration, our microscope or telescope or looking glass, whatever metaphor works for you, will be language. We'll be looking at what we can learn from language about the human mind, what language teaches us about the human mind, and we are going to specifically look at what linguistic diversity, crosslinguistic variation, and how languages frame the world and represent reality, can teach us about the mind.

This lecture series will focus on two topics, namely, the title is "Field Methods in Cognitive Linguistics." Now "field methods" is something that you need for studying any linguistic phenomenon in the field, whether it is syntax, phonology, or morphology. But we are going to be interested specifically in field semantics. So we are going to look specifically at methods for studying semantics in the field and in semantic typology. Those are the main topics of this series of lectures, and in fact we are going to focus on field semantics during lectures 2–4 and on semantic typology during lectures 5–10.

So we start with field semantics. What do I mean by field semantics? Basically, I'm going to try to introduce some methods and some theoretical background, epistemology, meaning the theory of how you get to know stuff in this particular field, methods and epistemology for what I call empirical semantics, which is methods for the empirical study of semantics, meaning the study of linguistic meaning on the basis of observation rather than interpretation. Semantic typology on the other hand is the crosslinguistic study of

All original audio-recordings and other supplementary material, such as any hand-outs and powerpoint presentations for the lecture series, have been made available online and are referenced via unique DOI numbers on the website www.figshare.com. They may be accessed via this QR code and the following dynamic link: https://doi.org/10.6084/m9.figshare.11419101

© JÜRGEN BOHNEMEYER, 2021 | DOI:10.1163/9789004362628_002

semantic categorization, so that's where we are looking at how different languages represent—frame—reality, the world, and what that teaches us about the human mind.

So the data for semantic typology primarily comes from semantic fieldwork. We generally have to go and collect primary data on how speakers of different languages frame a particular kind of stimulus, or a particular kind of object or action or spatial relation, and then we can do typological analysis. We can compare and try to find out what the languages of the world have in common in this respect and how they differ from one another.

What is that good for? It gives us a chance to sort out universals and variability in semantic representation. That in turn, as I will explain in a second, will allow us to map the nature-nurture divide in cognition, meaning it allows us to tell what aspects of human cognition are possibly innate and universal and what are learnt and culture-specific and therefore variable across human populations. But more than that, this kind of research also helps us understand how the language-cognition interface works, meaning how language might affect nonlinguistic thinking in different populations, different peoples, and conversely how cognition affects what is expressed in language. Finally, we can use semantic typology as a window into the syntax-semantics interface, the principles that govern the mapping between linguistic meaning and linguistic form, morpho-syntactic form, morphology, and syntax. And I'm going to illustrate that with examples shortly.

Now I want to contrast this program against what I call a traditional field methods course, which is the way a field methods course is taught in the places that have a tradition for teaching it. I don't know whether China has a tradition of teaching field methods. Russia does, the United States do, the Netherlands do actually, a few places in the world do. Now the way it's taught in the United States, including by me—it's, in my department, in my university, it's a two-semester course and we get a speaker of some indigenous language that the students don't speak and that I don't speak, and that student will be our consultant or our informant, the person from which we collect data just like real field workers collect data from the speakers of indigenous languages in the field.

So that kind of course is very much practice-oriented. It's very much focused on learning by doing which is important because fieldwork is ultimately not a theoretical endeavor. It's a practical endeavor; it's something that you do and therefore that you can learn only by practice. So that kind of practice you are not going to get in this course because you can not convey that in lectures because people won't learn it unless they actually get a chance to do it. This course is going to focus on the theoretical side of the fieldwork. It's going to

focus on methods and epistemology, and that can't quite compensate for the experience that you can get from taking a field methods course or more importantly, just going to the field and doing field work yourself. That's ultimately how anybody who has ever done fieldwork has learned it.

So once again, the first three lectures are going to focus on field semantics—on how we can gather, extract, collect data for semantic analysis. The remaining six lectures are going to focus on what we can do with crosslinguistic data on linguistic semantics.

The question is how does all of this relate to Cognitive Linguistics? My rationale for focusing on field semantics apart from the fact that that's my specialty—I do work on semantics and on semantic typology—is that of course linguistic meaning is one of the central areas of research in Cognitive Linguistics, and understanding the relationship between nonlinguistic cognition—thought, reasoning—and language is surely one of the goals of Cognitive Linguistics. Semantic typology is one important means of accomplishing this goal.

So I'm going to talk a little bit about field research, to enrich this concept a little bit, get a better sense of what we mean by field research. That's a concept that's used very widely across different academic disciplines. You have it throughout the humanities, but also outside the humanities in the life sciences, the geosciences, and even in opinion research, for example, marketing research. It always means something slightly different. Every discipline has its own particular take on field research, but what they all have in common is it is research *in situ*, meaning 'in place.' It's studying the phenomena in the place where they occur.

Let's take a look at what this means just in the various disciplines that are close to linguistics, starting with anthropology. For those who know the classic four-fields theory of anthropology, according to which anthropology comprises archeology, physical anthropology or paleoanthropology, cultural anthropology, and linguistic anthropology: each of these fields has its own approach to fieldwork.

For archaeologists and paleoanthropologists, of course, fieldwork means digging, going some place and excavating some ruins or some, you know, ancient human bones and so on and so forth. For cultural anthropologists, field work means participant observation, going to the community they wish to study and living with the people, immersing themselves in their ways, trying to become a functional member of the community and thereby learning what it takes to do so. Learning the practices of this culture, learning the procedural knowledge that one needs to have in order to function as a competent member of a cultural community. That's participant observation. And then in linguistic

anthropology, you will do that, but you will also use the field methods that structural linguists use, which is what we are going to be talking about in these lectures.

Now in sociology, fieldwork primarily means interviews. You have a questionnaire, and you go and interview a large sample of people, record their responses, and then do statistical analysis over it.

Finally, in linguistics, as a matter of fact, we do all of the above. There are at least four different approaches to field work within linguistics, starting with the linguistic anthropologists, who combine methods of cultural anthropology and methods of structural linguistics and often also methods of sociology. So they may, for example, run interviews on the basis of questionnaires, but also interact with a community for a long period of time in order to study the members in participant observation.

You then have the sociolinguists in the tradition that was mainly created by Bill Labov, in which people primarily rely on questionnaire data. Then you have the traditional dialectologists, an approach to the study of dialect diversification, the distribution of the different dialects of a language, which originates from 19th-century historical linguistics. And the traditional approach to dialect geography is basically for a single researcher to go from village to village and in each village identify what they call the most conservative speaker—which typically is an older, not well-educated, male. Then they record that speaker, have them pronounce a few words, say a few sentences, and so on.

Finally you have the structural linguists, and I'm including here the semanticists, meaning the people who study language as a semiotic system: its syntax, its morphology, its phonology, and its semantics. All of that I'm including under structural linguists. So what we do is, we mainly rely on three sources of data. We do, to some extent, and to an increasing extent nowadays, record spontaneous conversations as they occur without the intervention of the researcher. But that's often very hard to do in traditional communities, because it's not something that people necessarily welcome, and may be something that they find very disruptive.

Aside from that—and that is of course the type of research that is particularly important for Conversation Analysis, if you are interested in studying the strategies that people use to manage their conversations, which is something that you cannot do if you as the researcher are influencing what is happening. But aside from that, syntacticians, semanticists, and so on mostly rely on these two sources of evidence: recording what Nikolaus Himmelmann (1998) has called staged discourses, such as shoving a microphone into a speaker's face and asking them, "Ok, tell me a story." And of course they will tell you a traditional story just as they would to a child in the community or even an adult in

the community. So the story you are getting may be more or less conventional, traditional format that you did not have any influence on. But nevertheless, this is called "staged" because the reason this story is told at this particular moment in time is because you, the researcher, asked the speaker to tell you that story. So that's not a naturally occurring event.

Finally there's elicitation, and that's going to be our main focus of emphasis. What does elicitation mean? It means that you provide the speaker with a particular task and a stimulus to which they respond. The stimulus may be an utterance in their native language or in a contact language. It may be a picture that you ask them to describe; it may be an utterance in their own native language that you ask them to judge as to whether it's well-formed or not, and so on and so forth.

Now what exactly do I mean by "in situ"? This is actually a tricky question because "in situ" suggests that you are studying the phenomenon as it naturally occurs. But as I already mentioned, you have to generally assume that the observation interferes with the occurrence of the phenomenon. And this is something that the different disciplines that do fieldwork try to take into account in different ways. For example, the sociolinguists have long observed that the responses you get to a questionnaire actually depend on whether people, the respondents, perceive the interviewer as somebody who is a member of the community themselves or not.

There's a classic study by Peter Trudgill conducted in the early 1970s in northeastern England, where he was able to get people to respond in their local working class dialect pronunciations because Trudgill himself, despite the fact that he was a university professor, was using these kinds of pronunciations. So he was being perceived as a member of the local community, and without that people would not have responded in that way. They would have used Standard English pronunciation.

Now in structural linguistics, we have a very peculiar understanding of what it means to do field work, which has been influenced by a very peculiar history of this field of structural linguistics, which, as we are going to see, has been sort of meandering its way between the humanities and the social and behavioral sciences. Over the history of modern linguistics, which is basically a field that developed in the 20th century, you have a whole lot of idealizations, which have accumulated due to the heritage of linguistics, which mostly comes out of the humanities. As a result, it is often assumed that as long as you are collecting data from a speaker of the language, that is fieldwork. It doesn't actually matter where you collect the data or who this person is that you collect the data from, and so on. One speaker anywhere in the world, it may be in a hotel room and may be in a classroom—that counts as fieldwork. That is a very widespread

notion, and it's a problematic notion. Because of course there is pretty much no way that what is going to happen here really is a fair representation of what happens naturally in the community. But that was our original, naïve understanding of the meaning of "in situ."

Nowadays those of us who do field work are increasingly concerned with preserving linguistic diversity. Or maybe not linguistic diversity itself, because that is often a bar set too high already for what we can achieve. Nowadays we mostly struggle to preserve a record of the linguistic diversity that we find on the planet now, as this diversity is unfortunately rapidly diminishing and more and more languages become obsolete. They are no longer spoken by the members of the ancestral community.

When you're faced with that kind of situation, where you have very few full speakers left, and maybe all the full speakers that are left are in their seventies or eighties, maybe in fact the language is pretty much no longer used in the community. Maybe there are two elderly ladies and they live on opposite sides of town and they don't really have much contact and only when they get together once a year is the language still spoken. So then of course what exactly it means to do fieldwork under these conditions is a whole different ball game. It is a challenge that many of those of us who are interested in doing fieldwork have been facing.

As a matter of fact, it's also something that has been motivating a whole new generation of students to go into linguistics and learn the trade of documenting and describing endangered languages and going out and producing a record that might be of use for the study of the human mind, because every language is an invaluable data point in understanding the human mind, but also for the community of speakers, in case one day they want to try to revive the language.

Field semantics is the study of semantics in the field, under field conditions. The question I want to ask right now is why would you do that? Why are people interested in doing that? So field semantics is something that occurs for a number of different reasons. First of all, field semantics has its place in language description and language documentation. What I mean by that is basically: language description produces a scientific record of the language that linguists and cognitive scientists can use to study the grammar and lexicon of that language. What do I mean by "language documentation"? Language documentation is a scientific record of how the speakers use the language. It's a scientific record of the linguistic practices, and that's going to be important for a lot of researchers who want to study the language and the community of speakers: anthropologists, conversation analysts, people working in pragmatics, and so on and so forth.

But at the same time, the documentation ideally also would enable the speakers of the language or rather the members of the community—possibly at a point when the language has become obsolete—to revive it. Which is something that is actually happening in a number of cases around the world. Indigenous languages that were no longer spoken for a period of hundred years or more are now slowly coming back thanks to whatever records are in fact available. So one case in point is the Algonquian language of Martha's Vineyard, an island of coast of the Massachusetts, a language called Wampanoag, which wasn't spoken since the early 19th century, and nowadays members of the Wampanoag community, whose native language of course is English, are learning their ancestral language again.

Semantics has its place in language description and documentation, just as syntax and phonology have their place. But as a matter of fact, the people who typically go out to study and describe and document languages in the field primarily have the training to study syntax and phonology and not semantics. Why? Because our field, the field of linguistics, has had a bias towards syntax for the last fifty, sixty years, more or less since the Chomskyan turn in linguistics, and it has a bias toward phonology and morphology and syntax for an even longer term, namely pretty much since the inception of modern linguistics. So, as a result, people typically don't have the training they need to study semantics in the field.

Of course a result of that is that the semantic side of language descriptions often lacks. You have people confusing, let's say, perfective aspects for past tenses because they happen to be more familiar with past tenses or they're mistaking an evidential marker for a perfect aspect because they are more familiar with a perfect aspect. So in other words, Eurocentrism, essentially, or, more generally speaking, a bias towards whatever the researcher knows about their own language and related languages, is the result. That's one of the reasons why, when you are doing semantic typology, you have to usually go and collect the data yourself. Because existing language descriptions, grammars and lexicons, dictionaries that you take off the shelf, are not going to give you the answers for how particular ideas are expressed in the languages of the world, because that's not what the researcher who compiled the grammar looked at.

Next there is also a growing body of research on semantics in the field that is directly driven by theoretical goals. Meaning the people go and study semantics in the field because they have theoretically inspired questions. Such as, within Cognitive Linguistics, questions that are for example inspired by Anna Wierzbicka's "Natural Semantic Metalanguage" program. People want to know how certain proposed semantic primitives are expressed in different languages,

and whether they truly are primitives. There has been a lot of research inspired by the work of my colleague Len Talmy, who also gave lectures in this forum a few years ago. Research inspired by Ray Jackendoff's Conceptual Semantics. I've done some work on the basis of that myself.

Outside Cognitive Linguistics, in formal semantics, there is also a growing interest in fieldwork and a growing community of field working semanticists. One of the forums that have developed over the last ten years is a series of conferences called "Semantics of Underrepresented Languages in the Americas," which focuses on formal semantics in Native American languages. We had the most recent one at Cornell University this May.

Finally, there is research in field semantics driven by semantic typology, which is my own main line of research, the cross-linguistic study of linguistic categorization. For example, color term systems. Which is something that we'll talk about more extensively in the fifth lecture. So what you have here, you have on the one hand languages such as English or, I believe, Mandarin which, at the level of basic words—meaning words that are morphologically simple and they are used by every speaker of the language; they are not loan words, at least not easily recognizable as loan words—at that level of so called "basic color terms," a distinction between blue and green. On the other hand, there are many languages that have a basic-level distinction between dark blue and light blue. That is not the case in English, but it is the case in Russian and Turkish and Persian and many other languages of that particular area.

And then there are many languages, like Yucatec Maya, the indigenous language that I've been studying for twenty years, which has a single word for green and blue, a word that means something like 'grue'—'green and blue.' Now this is not 'colorful.' It doesn't cover orange, or yellow, or brown, it's just green and blue. The question is, given this kind of cross-linguistic variation, are there any constraints on this? Are there principles that the languages of world share that govern the structure of color terminologies? And if so, what does that suggest about the role of cognition, the nonlinguistic perception and categorization of color in language, and, conversely, what does this suggest about the role of linguistic color terms in cognition? For example, is it the case that Yucatec speakers find green and blue more similar to one another than English speakers, and English speakers find light blue and dark blue more similar to one another than Russian speakers? And the answer by the way to the last question is yes, they do.

As semantic typologists, we map the distribution over different categorization strategies across the languages of the world—which is what typologists basically do. They map the distribution of linguistic phenomena, phenomena

of syntax, phonology, or semantics. So here you have a map from the World Atlas of Language Structures that shows the distribution of green-blue distinctions across the languages of the world. The red dots are languages that have distinct basic color terms for red and blue, whereas the yellow circles are languages that have a single term for green and blue, and then the other stuff—that's other stuff.

As typologists, we are trying to formulate generalizations over this distribution. We are trying to find principles that govern this distribution. In the case of color terminologies, originally Berlin and Kay (1969) proposed a set of generalizations, and the latest manifestation of these generalizations were proposed by Paul Kay and Luisa Maffi in the late 1990s. What this basically means is that if a language has basic color term systems, then the simplest way of doing that is you just have two terms, one that covers all the light-warm colors and one that covers all the dark-cool colors. And then if you have three terms you have to break out white. You have to have a distinct term for white. Then there are different routes that you can take to by and by break up the so-called multifocal colors into single focal colors until you reach this inventory of six monofocal basic color terms. We'll talk more about that later.

That was a quick road map, a very simplified road map of what semantic typologists do. Now the question is, *why* do they do that? What's the goal? What's the big idea? Why is this interesting? Why do we do this kind of thing? Why do we want to study semantic categorization across languages? The answer, in a nutshell—well, this is sort of the bumper sticker answer, meaning it's very simplified version of the answer—is that studying linguistic categorization allows us to distinguish between what is variable in terms of semantic categorization and what is non-variable, and that in turn allows us, a pretty good guess at least at what's culturally transmitted learned knowledge and what's potentially innate knowledge.

I wrote a white paper for the National Science Foundation, basically telling them that they need to fund this kind of research. Which of course you know you could argue I have some vested interest in. I called this "mapping the nature-nurture divide in cognition," where nurture of course is a metaphor for culture. So what we are trying to do is find out where in the human mind in cognition the dividing line between learned and innate knowledge runs because we expect innate knowledge to be invariable across human populations and learned knowledge to be culture specific, and therefore, surface in the form of cross-linguistic variation.

In the process of doing this, we also have to clarify the relation between nonlinguistic cognition—internal thought and reasoning, memory and so on—and language. How does cognition influence language? How does language

influence cognition? Which of course is the Whorfian question, the linguistic Relativity Hypothesis. And as I already said, in the process we also try to contribute to theories of the syntax-semantics interface and I'm going to illustrate that in a moment.

Since we'll be talking a lot about knowledge and representation and cognition, let me give you, in a nutshell, some very naïve, heavily simplified characterizations—I'm even not going to call them definitions—of what I mean by all these things. One thing that's important to me to clarify is: all of the stuff is very much in flux for me. So, my thinking about these issues constantly evolves, what you are seeing here has been the result of twenty years of thinking of evolution, and if I'm still alive another twenty years from now then, I hope that at best I'm going to be laughing at this.

So what do I mean by representation? Basically one entity or state of affairs standing for another entity, conveying some sort of information about another entity or state of affairs. A representation in some sense is always a sign. It is not necessarily symbolic, and it doesn't have to be a conventional sign. But if Charles Saunders Peirce's classic classification of signs is correct, then every representation can relate to that which it represents in exactly three possible ways: As a symbol, meaning by way of convention, or a learned connection culturally transmitted. Or as an index, meaning something that stands in a causal or a contiguity relationship with that which it stands for. Imagine footsteps on a beach as a representation of the person or animal who walks there, or smoke as a representation of fire, or storm clouds as a representation of a coming storm system, and so on. This morning I stepped on the scale in my bathroom, and it gave me a reading, which then gave me pause. But that reading is an indexical representation of my body weight. And my body weight is going up thanks to the great food in this country.

You have iconic representations which are representations constituted by similarity between the representation and that which it stands for, such as a picture, a map, a painting, a film, the video recording of these lectures, and so on.

So I have a very, very broad and wide sense of "representation," and I certainly don't mean by representation just symbolic representation, which is important because I think that this equation of representation with symbolic representation has led to a lot of confusion in cognitive science.

Next, what do I mean by "meaning"? That is the information that the representation conveys about that which it represents. That sounds like it's the easiest, most straightforward, trivial thing in the world, but as a matter of fact, the hard part is explaining what the heck we mean by "information," and that's what semantic theory tries to do.

What is cognition? This is probably the most difficult one of the four. I'm going to take a roundabout route by first trying to define the notion of a cognitive agent, a cognizer if you will, which can be a human being, it can be an animal, it can be a robot. It doesn't matter. A cognitive agent is a biological or artificial entity that is capable of generating and processing representations of its own internal states and of its environment. We certainly expect cognitive agents to some extent to be able to act on these representations and therefore possibly change their environment. Cognition then is both the process of generating and processing these representations and also the ability to do so, to generate and process these representations.

Finally, what about culture? Culture is the sharing of representations between cognitive agents. What do I mean by that? It's the sharing of knowledge. There are two important flavors of knowledge that are distinguished in cognitive science, called "procedural" and "declarative" knowledge. Declarative knowledge is knowledge of facts that you can convey linguistically, in terms of linguistically encoded propositions, such as, let's say, the kind of knowledge that you learn in history classes in school. Procedural knowledge is knowledge how to do something, which you basically can only learn by doing. Remember what I said about fieldwork. Since ultimately fieldwork is a practical enterprise, you have to actually do it in order to learn it. Same as the knowledge you need to ride a bike. That's not knowledge you can get from somebody explaining to you what you need to do to ride a bike. That explanation may be perfectly sound but it does not by itself enable you to ride a bike. You simply have to go and try it and learn it yourself. So that's procedural knowledge.

The social sharing of both procedural and declarative knowledge is culture, but also the social sharing of practices, meaning it's not necessarily just the knowledge of what people do, it's the doing itself, the habitual doing, the styles of action, the styles of speaking, the styles of walking, of comporting yourself, dressing, responding to particular gestures, but also cognitive styles, styles of thinking, of reasoning, which are shared among the members of a community. You're going to say, oh, wait a minute, how is that possible? How is it possible for the members of a society to share styles of reasoning? How the heck are they going to do that? Are they going to be telepathic or what? And the answer is, we can infer styles of reasoning from the actions and behavior of other members of our community. We can infer those to some extent from their nonlinguistic actions, but in particular, we can infer their styles of thinking from their gestures and from their use of language. That is going to be one of the bigger topics of some of the lectures that are ahead of us. As I said, these are just sort of thumbnail sketches. You get a longer story of what I have in mind when I talk about some of these things in the appendix of the handout.

So any kind of knowledge is going to be either innate or learned. If it is learned, it is going to be either learned from individual experience, such as the famous stereotypical hot plate on a stove that supposedly every child has to touch in order to learn that that's not a good idea. That's learning from the individual experience. And then there's knowledge that is culturally transmitted. This is a huge chunk of everything that we know. Notice that cultural transmission of knowledge is what's happening right here right now. I'm transmitting knowledge to you, and you are also transmitting knowledge to me, because, you know, I'm looking at you guys, I'm studying you, I'm trying to make some inferences about audiences in these kinds of lectures in China, trying to learn about that. And probably what I'm coming up with is a lot of very invalid generalizations, partly because the sample is too small.

But, be that as it may, all of us do this all the time, and a lot of what we know—mathematics, music, the law, economics, this huge body of human knowledge—is not innate. None of it is innate. All of it is cultural, and therefore we possess this knowledge neither because it's in our DNA, in our genes, nor because we figured it out ourselves, but rather because we learned it from our parents, our teachers, our friends and peers, and so on and so forth, in other words, through cultural transmission.

This is a very abridged and simplified rationale. Anything that you find, any aspects of human knowledge that are variable across human populations, are generally assumed not to be innate. The underlying rationale for this argument, for this analysis, is that there are no established cognitive differences across human populations. We know there are some genetic differences across human populations, although these are apparently very superficial. That doesn't mean that people have not given them enormous meaning, usually with very negative consequences. But genetically, biologically, these differences that express themselves in pigmentation and so on and so forth are very shallow, and, as far as we know, there is no evidence that suggests there are any cognitive differences that are genetically transmitted. So in other words, we have reason to assume that whenever we observe culture-specific differences in human behavior, the relevant behavior has to be learned, it cannot be innate—because then it wouldn't be variable.

On the other hand, if we observe an aspect of human behavior across populations that's not variable, then one possible explanation for that uniformity is innateness—but it's not the only explanation. Talking just about linguistic behavior, and about the structures of language that we observe—for example, the structure of semantic representations that we semantic typologists study—we observe that if there is an absence of variation, maybe one reason for that is because of the general design features of language. This is the very highest-ranking design features, such as the fact that there is the so

called "double articulation" in the sound structure of languages. So you have meaning-bearing and meaning-distinguishing elements, i.e., morphemes and phonemes, and these things are not identical. There will be properties of language that just follow from these general design features. So the properties that are entailed by the general design features do not have to be innate even if some of the design features are innate.

Finally, there is the possibility of monogenesis and inheritance from the common ancestor, the putative hypothetical common ancestor of all surviving human languages, if there was such a common ancestor. Since by some accounts, human language is a very recent phenomenon and has only been around for a few ten thousands of years, we may be able to trace back all extant human languages to a single ancestor. Not in actual fact—we have not been able to do it. But some people think it is possible in principle, and therefore, a lot of the traits that we observe across the human languages that exist today occur just because they have been handed down through the lineages that have evolved from a common ancestor.

Of course there are several problems with this argument. First of all, monogenesis isn't assured. Monogenesis isn't a foregone conclusion, I think. In addition there are new languages, which don't descend from the common ancestor, Pidgin and Creole languages. So if those languages show the same feature, then the argument wouldn't carry over to that. But finally, aside from all these other possibilities, there is the possibility of innateness. So one possible explanation for why certain features are shared across the human languages might be innateness.

You could also take a more dynamic view of cognition in terms of categorization or conceptualization, the dynamic processes involved in forming concepts. And if you do that, then maybe cognition doesn't need any kind of innate categories, any kind of genetically encoded knowledge that we have from birth. All it really needs is something like an innate categorization engine, which forms concepts in predictable ways in response to environmental stimuli. And of course part of those environmental stimuli are going to be the cultural and linguistic data that the child learns from other members of the community. So this opposition between innate and learned knowledge, this is something that most people in cognitive science have been taking for granted since the inception of cognitive science. Because the basic idea of cognitive science is that the mind does not start as a blank slate, *tabula rasa* as they say in Latin, but rather there's some sort of presetting that allows children to learn other stuff. But I believe that that's not necessarily true.

And secondly, the idea that human languages may share properties simply because these properties descend from a common ancestor is a consideration. This is a hot new topic of research that has been generating a lot of very

interesting and provocative studies in recent years, thanks to the new availability of methods of phylogenetic statistics, originally developed in evolutionary biology, which we can now use to compare languages, the features of different languages, and come up with likely scenarios for the evolution of these features. We'll talk a little bit more about that in lecture 7, the lecture on quantitative methods in semantic typology. I'm skeptical about this line of research for several reasons, including the fact that I'm not sold on monogenesis. I'm not convinced that all extant human languages must have come from a single common ancestor. One reason to be skeptical about this is the fact that there is a lot of circumstantial evidence to the effect that ancient humans, meaning the Neanderthals and so on and so forth, already has some form of language. For example just last week there was an article in the journal Science that pointed out that a lot of the drawings of hand shapes that you find in the caves of Europe probably were made not by modern humans but by the Neanderthals. Because the hand does not quite have the shape of the modern human hand, and because it also seems to be too old. I mean some of these drawings seem to be simply too old to have been made by modern humans. Which suggests that the Neanderthals already had symbolic art as they call it. They most certainly controlled fire. They had complex hunting techniques that it's very hard to imagine how they could have shared them culturally among one another without some form of language.

And if they had some form of language, that means that modern humans—who most assuredly are all related to a single of group of common ancestors that originated in Africa—could have easily picked up language from ancient humans in different parts of the world separately, which essentially would mean that human language as we know it may have had multiple origins, although all of these ultimately would develop from whatever proto-language was spoken by ancient humans. That's all speculation. All I'm trying to say is that monogenesis is not necessarily something that we should take for granted.

Finally, there was a very important paper by Nick Evans and Stephen Levinson in Brain and Behavioral Sciences in 2009, where they point out something that most people in cognitive science are not aware of, which is there are very few, if any, strong language universals, meaning very few principles that hold for all languages without exceptions. So according to Evans and Levinson, the emphasis on language universals has been strongly overblown. I agree with that whole-heartedly. But I think there is another side of the picture which Evans and Levinson omit, which is that there *are* a lot of striking similarities across the languages of the world in terms of the meanings that are expressed. Even if they are not exceptionless, even if to any possible generalization that

you can come up with, there will be one or two, ten, or twenty languages that don't conform to that generalization, you still have to explain why it is that thousands of other languages do share a certain feature. That's a very important task of semantic typologists.

These are three different versions of what I called the "big picture," which you got in the blueprint a moment ago: the relation between nature and culture on the one hand, and the relation between language and nonlinguistic cognition on the other. The first version of the big picture was assumed by people like Frans Boaz, Edward Sapir, and Benjamin Lee Whorf. This is pretty much what people on both sides of the Atlantic took for granted until after World War II, until the advent of the cognitive sciences. According to this picture, most of cognition and language, in fact probably all except for trivial parts, would be learnt, would not be innate, would therefore not be determined by biology, and language would play a very important part in shaping the categories of thought, of internal cognition.

Then came the Cognitive Revolution, the Cognitive Turn, the rise of the cognitive sciences and, in linguistics, the rise of generative grammar, Chomskyan linguistics. With that new program, people have emphasized very much the inverse of this picture. A version according to which almost everything interesting about cognition is innate, and the only parts that are learned are trivial and uninteresting. And language has almost no role in internal cognition except a more or less trivial function in defining certain constraints on what can be expressed and therefore what's useful to entertain in the way of thought. By and large what is expressed in language simply is an external representation of internal thought. And that is the version of the "big picture" that has been championed and advocated by Noam Chomsky, Jerry Fodor, Lila Gleitman and many others. I would say it is still the case in the United States today, if there is a dominant paradigm—and there *is* a dominant paradigm; I'm not sure that is still true of Europe or Asia, but in the United States it's still true—and this would be it.

Then there are the Neo-Whorfians, who have been tearing, drilling holes, into this version of the picture, and basically have been trying to demolish it for the last twenty years, such as Stephen Levinson, Melissa Bowman, and Dan Slobin. So there is a role, possibly, for biology, and cognition in language that is a little more substantial than what Whorf and his contemporaries assume. But on the other hand, it is definitely true that a very significant part of cognition is cultural, is learned and socially shared. And language plays an important role in that because it is a very important means of transmitting cultural knowledge. Not the only one, but a very important one, possibly the most powerful one that humans have at their disposal.

Now I want to give some brief examples to illustrate how we can get from semantic typology to research about the mind and cognition. And the first example is the classic research on basic color terms by Berlin and Kay. Berlin and Kay went and gave a standardized set of color chips to speakers of 20 different languages and asked them, "How do you call this stuff?" "What colors are the best instances of the basic color terms?" They found that there were some principles that were shared across the speakers of these different languages, including what chips serve as the best instances, the best exemplars of the colors. Eleanor Rosch then developed the theory of Prototype Categorization in cognition, which was heavily inspired by Berlin and Kay's work. Kay and McDaniel proposed in 1978 the Neural Substrate Hypothesis, which tries to attribute the findings—the generalizations—of Berlin and Kay directly to properties of the neurophysiological processing of color perception. That in the meantime has turned out to no longer be tenable. It was based on data from color perception research in the 1960s that was later shown to be invalid. But still, sooner or later, people are going to try to go there again and build a bridge all the way from the observable linguistic data to the underlying neurophysiological substrate.

In 1984, Paul Kay and Willett Kempton demonstrated a weak Whorfian effect of color categorization. In other words, they were able to show that speakers of Tarahumara, which is a language that, like Yucatec, has a single word for green and blue, do indeed, if you run a similarity judgment task, find shades of green and blue more similar to one another than English speakers do. More recently, Terry Regier and colleagues replicated these experiments, and they were able to show something very intriguing, which is that these Whorfian effects, these language-specific effects, only manifest themselves in the left hemisphere. So, if you shut your left eye, and you only get the stimuli administered through your right eye, which is connected to your left brain hemisphere, which is where language is situated—most of language is lateralized in most people on the left side—you get the language specific effect. But if you do it through the left eye, with the right eye closed, you get no effect. Very intriguing.

In the late 1990s, it was finally demonstrated that not all languages actually have basic color terms, something that had been heavily disputed for decades, partly thanks to research by Stephen Levinson, who showed that the language of Rossel Island in Melanesia does not have basic color terms. In fact many languages don't have them. So that was acknowledged by Paul Kay and Luisa Maffi in 1999. In the meantime, there has been a new strand of research around Debi Roberson from the University of Essex in England, suggesting that maybe language indeed plays a strong role in color categorization after all. In response,

Terry Regier and Paul Kay and colleagues developed new statistical methods of testing the similarity among color categorizations in the World Color Survey, which was an analysis of the color term systems of over 100 languages conducted originally in the 1970s. Regier and colleagues found that, yes, there are universal principles of lexicalization of color, contrary to Debi Roberson's and colleagues' claims. So this is all ongoing research, all very controversial, but you see how you can get from cross-linguistic semantics directly to cognition and even neurocognition.

Another domain is spatial frames of reference. Until the 1970s, it was assumed that all human populations have a universal innate bias towards using relative, egocentric frames of reference in small-scale space. So they would all, let's say, represent the position of people in the audience in front of them in terms of 'left' and 'right', or in terms of 'front' and 'back', relative to where they, the speakers, are located. Then it began to emerge in the 1970s that a lot of Australian aboriginal populations represent everything in terms of cardinal directions linguistically. And not long afterwards, John Haviland was able to show that that also extends to gestural representations. So gestural, iconic representations of real-world events in story narratives turned out to be oriented absolutely so as to preserve the cardinal directions of the represented event in the gesture. And that then led Stephen Levinson and colleagues to do a lot of work on indigenous languages of Mexico, and Australia, and Namibia and other parts of the world and demonstrate that preferences for particular strategies of spatial orientation in discourse, in language, align with preferences in speakers of these languages in how they memorize spatial information. They use the same frames of reference, the same strategies in language and in recall memory. They also demonstrated that speakers of languages that have a bias in favor of absolute orientation, geocentric orientation, tend to be better dead-reckoners, meaning they are much better than speakers of languages that prefer egocentric or intrinsic orientation in terms of keeping track of how their bearings change as they move through space.

Levinson and Pederson and colleagues explain this in terms of a language-on-thought effect, a Whorfian effect of language on internal cognition. Peggie Li and Lila Gleitman and colleagues pushed back against that, arguing instead, no, it's not language influencing internal cognition; rather it's culture influencing both language and internal cognition. Daniel Haun in 2006 with a team of researchers show that non-human primates, chimps, bonobos, gorillas, and so forth have a preference for geocentric cognition. Which is very interesting, given that until a few decades ago, it was assumed that egocentric cognition was innate in modern humans. Which would mean that most likely, if that

were true—which nowadays very few people still believe—then preferences for spatial orientation would have had to be encoded genetically at least twice in the hominid lineage.

Most recently, Daniel Haun and colleagues demonstrated that it's actually difficult to train children who have been growing up using predominantly geocentric frames of reference to use egocentric frames of reference instead and vice versa. So they worked with a group of around five year old kids in Namibia, who are growing up in a community that prefers geocentric orientation, and tried to teach them relative terms and to solve a task of nonlinguistic reasoning in relative terms, and conversely they tried to teach a group of Dutch kids who grow up with an egocentric bias to use a geocentric strategy to solve the task. In both cases, the kids perform below chance.

The final example I wanted to give comes from research on motion representations, which was started a long time ago with Len Talmy's doctoral research on Atsugewi, a Californian indigenous language, a presumed member of the hypothetical Hokan stock.

Len noticed that Atsugewi classifies motion in terms of properties of the figure, the moving objects. They have different motion verbs for, let's say, round objects and flat objects and so on and so forth. He then went on to compare that to English, which often conflates the manner in which an object moves—whether it's running or walking or rolling or swimming and so on and so forth,—to languages such as Spanish, Turkish, or Japanese, which conflate the path of the motion in the main verbs. So they talk about entering, ascending, descending and so on and so forth. And then there are languages such as Mandarin, which use verbs for both of these things.

In the 1990s, Dan Slobin began to show that the syntactic type of a language in terms of Talmy's typology has an influence on how speakers of the particular languages prepare their nonlinguistic thoughts when they are trying to talk about motion events. So the speakers of languages that conflate manner of motion talk in the verb a lot more about manner of motion than the speakers of languages that conflate the path of the motion event in the verb.

Since then, a lot of different people including myself with a team of researchers have been trying to test whether this has an influence on the nonlinguistic categorization of motion events. There's also a syntactic side to Talmy's typology, although Talmy talked about it primarily in terms of lexicalization, which has been examined by a lot of different people, including Bhuvana Narasimhan. In more recent research, a bunch of colleagues and myself showed that the different linguistic types for motion event framing actually impose different constraints on the events that get represented as simple events, if you will, in these languages. So whereas an English speaker will take it for granted that

motion from point A to point B is a single simple event, whatever that means, because in English you can represent that as a single simple clause, in Yucatec Maya and many other languages you have to use a minimum of two clauses. You have to say "It left point A and then it reached point B." You can't just say "It went from point A to point B."

There has been some new research by Dan Slobin and Melissa Bowerman and colleagues, showing that children are sensitive to these different typological patterns of motion framing at a very young age. And of course language acquisition research or child language developmental research is always an important companion to semantic typology, in that if we assume that there are certain universal principles underlying the linguistic representation in a given domain, then we should expect that children, all children, regardless of what language they learn, start out with these universal principles and only diverge in language-specific ways relatively later. Whereas what these people found is that, no, even at the very early stages that they looked at, the kids already were behaving language-specifically, which is evidence against these universal principles in this case.

So that's it. Quick summary, field research is the data gathering "in situ", meaning wherever the phenomena occur. Field research in linguistics means data gathering from the speakers of the language. Field semantics is the gathering of data for semantic research in the field from speakers of the language for the purposes of language documentation, description, and theoretically driven research, and of course also for the purposes of semantic typology.

Semantic typology is the crosslinguistic study of semantic categorization, meaning the search for universals and variation in how different languages represent reality. Representation means one entity or state of affairs standing for another entity. Meaning is the content of the representation. Cognition is the generation and processing of these representations in a cognitive agent. Culture is the social sharing and transmission of knowledge and practices. And crosslinguistic variation can serve as indirect evidence of cultural transmission for the reasons that we discussed. Alright, now, that's it. Thanks so far!

LECTURE 2

Field Semantics: Studying Meaning without Native Speaker Intuitions

Thank you, welcome back. This is the first in a series of three lectures that are dedicated to field semantics, meaning the methods and epistemology for studying linguistic meaning in the field from the native speakers.

This morning I said that meaning can be understood as the content of a representation. A representation being a state of affairs or an entity that stands for another state of affairs or entity. That information that the representation conveys about that which it stands for, that's the meaning. There are two ways of interpreting meaning when it comes to signs and the way human beings use them, produce them, and understand them. On the one hand, we can interpret meaning in terms of referential content—a linguistic utterance as a representation of some state of affairs in the word. And on the other hand, we can think of meaning in terms of representational content, meaning an utterance as being an external expression of an internal mental representation, a thought in the mind of the speaker.

So if I say *It's hot and humid today in Beijing*, then on the one hand, I'm referring to a state of affairs in the world. I'm making a statement about the weather in Beijing, in a particular place at a particular time. And on the other hand, I'm telling you what I'm thinking about that particular state of affairs in the world. Different linguists and cognitive scientists tend to emphasize one aspect of meaning or another. These lectures are dedicated to Cognitive Linguistics. Cognitive Linguists tend to champion the representational perspective, meaning the view of meaning in terms of mental representation, and not so much the view of meaning in terms of reference.

Nevertheless, I'm going to argue that if we want to study semantics empirically from the speakers without relying on our own native speaker intuitions, we have to have some handle on the referential content of linguistic utterances.

All original audio-recordings and other supplementary material, such as any hand-outs and powerpoint presentations for the lecture series, have been made available online and are referenced via unique DOI numbers on the website www.figshare.com. They may be accessed via this QR code and the following dynamic link: https://doi.org/10.6084/m9.figshare.11419110

© JÜRGEN BOHNEMEYER, 2021 | DOI:10.1163/9789004362628_003

Why is that? Fundamentally, as field workers, we are not mind readers. Even if we believe that the meaning is ultimately in the minds of the speakers, how do we get to these meanings? How do we get to the internal mental representations of the speaker? Well, we can't directly share them. We generally speaking aren't very good at telepathy. We are really lousy at telepathy, which is why we use language in the first place. So we have to reconstruct what's in the mind of the speaker through what we can observe them referring to in the real world. That's basically the task. And my task in these lectures is going to be to flesh that out a little bit and try to make that a little more concrete.

I'm going to start with the distinction between "sense" and "reference," or "Sinn" and "Bedeutung" in the German original, which was introduced by the philosopher Gottlob Frege more than a hundred years ago. The original example that Frege gave is quite a bit involved. It goes like this: Imagine you are taking a triangle, and you connect each angle to the center of the edge opposite that angle. So that gives you three lines: A, B and, C. And it so happens as a property of geometry that regardless of what the triangle is that you start from, these three lines have to intersect in the same point. So that means the point in which A and B intersect is also the point in which C and B intersect, which is also the point in which A and C intersect.

So that means that the description *the intersection of A and B* and the description *the intersection of A and C* and the description *the intersection of C and B* all refer to the same point. They all have the same referent. But of course in every other respect they are semantically completely different. You wouldn't actually know that they happen to have the same reference necessarily without knowing something about the geometry of triangles.

Like I said, this is not necessarily the easiest example, so here are some easier examples. If I say to you: *Hey, your dog just bit me*. Then I'm using the word *dog* and I'm using it to refer to the particular animal that I'm identifying as *your dog*, and that's the referent in this particular usage. But the sense, the concept in my mind, is a classification, a categorization of possible referents that allows me to identify dogs and distinguish dogs from other animals and other stuff in the world like, let's say, a lectern or a bag.

So basically the conceptual content, the sense of the word *dog* somehow has to spell out criteria that allow competent speakers to distinguish possible referents of this word from those of other words. So I could say instead *Your pet just bit me* or *Your Chihuahua just bit me* or *Your mutt just bit me*. In that case, I'm using different words, all of which have different senses. So I'm making a different conceptual categorization. I'm categorizing the referent in a different way, but it's always the same referent. On the other hand, let's say, a different conversation, again I'm talking to you but now I'm saying: *Hey, Sally's dog just*

bit me. Now I'm talking about a different dog. So now I'm using the word *dog* again, with the same conceptual classification but a different referent.

Now this sentence does not make sense for all native speakers of English, but it does apparently to some:

(1) *Sam's not a dog; she is a bitch*.

That makes sense if you have a polysemous representation of the word *dog* in your mental lexicon such that it has one meaning which refers to the species that we call *dogs* and another in which it refers more narrowly to the male members of that species. So it has a separate sense 'male dog,' and under that sense, (1) makes sense. So, that means now I'm using the same word but with a different sense. That's the phenomenon of polysemy, which we will talk about some more.

So that's the distinction between sense and reference. Now there's a related distinction which was introduced by a student of Frege's, Rudolf Carnap, a little later, and that's the distinction between extension and intension. Extension is a useful term, which we use a lot in semantic typology. It's the set of possible referents of a term. So if the word that you are looking at is *dog*, then the extension of that would be all the possible critters that this word could be applied to. Of course if there are metaphorical uses of *dog*, let's say, for insulting somebody, or for saying well that a person is actually being treated poorly by somebody else, something like that, those would also all be included in the extension of *dog*.

On the other hand you have the intension. Now the intension is Carnap's equivalent of sense. It's the criteria that allow people to identify possible referents, which is the conceptual content in Frege's sense. Except Carnap gave a more formal account, which in the modern tradition of formal semantics has been interpreted as a function from possible worlds to referents. So you take a word and you put it in different contexts. You evaluate it with respect to different possible worlds and the function assigns a set of possible referents in each world, let's say one world there's only one dog, one world there are a hundred million different species of dogs, one world that is more or less like ours, and so on and so forth. This is not a very useful notion for the purposes of semantic typology. It's a notion that has been used to sort of gloss over the issue of sense in formal semantics. Formal semanticists aren't very comfortable talking about sense because sense is conceptual content. Formal semanticists are generally not very comfortable with conceptual content, since they don't have a whole lot to say about that, something we will talk a little more about shortly.

So the concept of intension as supposed to sense is not very useful for semantic typology, but the concept of extension, the set of possible referents is, for reasons that will become clear as we go along. And finally there is also the concept of denotation, which is again a notion that comes out of formal semantics, and it talks about the class of possible referents, what these have in common. Again that's not a necessarily particularly useful notion for our purposes.

Now here is a traditional model for thinking about the relationship between sense and reference, which goes back all the way to Aristotle. Aristotle originally proposed that in one of his books, which is called *Peri Hermeneias* in Greek, in the Latin translation, *De Interpretation*, "On Interpretation," something that I will get back to. Much more recently, this idea of the semiotic triangle was revived by Ogden and Richards in one of the classic books on semiotics.

The idea is basically that symbols only associate to referents—such as words having referents, for example the word *cat* referring to a particular animal—via some conceptual content in the mind of the speaker. So, you cannot get—that's the idea of the semiotic triangle—you cannot get from the symbol to the referent without some sense mediating between it. This is an idea that's been hugely controversial in modern philosophy. A lot of smart arguments have been advanced to the effect that this isn't always true, but it is certainly true in some cases.

So that means there are two fundamental axes to meaning, which we've already introduced: meaning as a relation between linguistic expressions and mental content, that's meaning as representation; and meaning as a relation between linguistic expressions and referents in reality or maybe in imagined reality, and that's the view of meaning as reference.

Now, meaning as mental representation is an approach that's been championed by what have been called variously internalist semanticists or representational semanticists, conceptual semanticists, cognitive semanticists, and so on and so forth. These are people who work primarily on phenomena that can be explained with respect to the properties of the underlying mental representations or cognitive representations in the mind of the speaker, semantic properties that can be explained with reference to cognition, such as lexicalization. This morning I mentioned Berlin and Kay's classic study on color terminologies—that's a problem of lexicalization. Lexicalization is the availability of words or constructions, linguistic expressions for concepts, a relation between concepts and linguistic expressions in particular languages. And of course in the case of Berlin and Kay's research, the question is what words do different languages have available in the mental lexicon of the speaker to refer to color senses. So that's the problem of lexicalization.

Semantic transfer, metaphor and metonymy, is a huge topic in Cognitive Linguistics. Formal semanticists have nothing to say about this, because it's not a problem that can be explained in terms of reference to the world. It's entirely in the mind of the speaker, entirely a cognitive problem. Strongly related, polysemy, the multitude of senses that you encounter in a single word and the relationship between these senses. Semantic change, the evolution of the family of senses of a word over time, and how we explain these changes. Prototypicality effects, i.e., distinctions in terms of the extension, a structure in the extension of an expression such that some referents are more prominent, more salient, better examples, exemplars of the category, than other members. This morning I mentioned Eleanor Rosch's work on prototype theory in categorization in psychology, which in turn was inspired by Berlin and Kay's work on lexicalization.

Taxonomies, the structure of taxonomies, families of concepts, such as, let's say, from life form you go to animal and on to mammal, fish, bird, insect, something like that. And from mammal you go to dog and cat and horse. And from dog you go to dachshund and golden retriever. You get the general idea. So that's the classification of concepts in terms of relations of superordinate and subordinate concepts, that's a taxonomy.

And metonymies, meaning part-whole structures in the linguistic expression. All of these are problems that formal semanticists have very little or nothing to say about and that have been given very insightful treatments in the cognitive or conceptual, internalist traditions. And here are some of the main people in cognitive approaches to semantics: Manfred Bierwisch, Ray Jackendoff, George Lakoff, Len Talmy, Anna Wierzbicka. Lakoff and Talmy of course have been prior speakers in this forum.

On the other hand, meaning as reference, that's the approach that's been championed in formal or compositional semantics. Let me by the way just insert here that the term "formal semantics" is often misunderstood. Formal semantics does not necessarily mean you are using a formal meta-language. Formal semantics means you are talking about meaning as contributed by the form of language—the morphosyntactic form. So we are talking about the semantic impact of syntax, by and large. That's why formal semantics is compositional semantics, according to the Principle of Compositionality, which was also introduced by Frege.

Formal semanticists have looked at problems such as scope ambiguity. And quantification has been a staple in formal semantics since its inception, because the standard formal meta-language is predicate calculus, which was developed by Frege among others in order to deal with the problem of quantification in logic. So you have a sentence such as *Every linguist has watched an*

episode of Red Dwarf, and that has two interpretations. One according to which there is a single specific episode that has been watched by every linguist, and another under which every linguist has watched at least one episode of this show, but not everyone has necessarily watched the same episode. Accounting for these different interpretations has been one of the issues that formal semanticists have worked on.

Entailment patterns is another. If you say (2), that entails (3):

(2) *Sally was watching Red Dwarf.*
(3) *Sally watched Red Dwarf.*

It's not possible for (2) to be true without (3) also being true. But if you say (4), (5) isn't entailed—it doesn't necessarily follow that Sally watched an episode of Red Dwarf.

(4) *Sally was watching an episode of Red Dwarf.*
(5) *Sally watched an episode of Red Dwarf.*

Why not? Because Sally might have been watching an episode of Red Dwarf when the fuse blew and the TV went dark and she couldn't finish the episode, so she never found out what happened in the end. Red Dwarf is a British Sci-fi sitcom of the 1980s for those who don't know it.

Implicatures and presuppositions. If you say *Floyd has watched some episodes of Red Dwarf*, that strongly suggests, but does not entail, that he didn't watch *all* episodes of Red Dwarf. In fact it is entirely possible, I mean, if it is in fact true that Floyd watched all the episodes of Red Dwarf, then it necessarily also follows that he watched some of them. So the phenomenon here is a scalar implicature, an implicature to the most informative interpretation of the utterance, according to the theory of generalized conversational implicatures that was developed by the philosopher Grice and published in the 1970s, although he had been working on it for a much longer period of time.

And then syntactic compatibility issue such as *Sally watched Red Dwarf for 20 minutes* versus *Sally watched the episode of Red Dwarf in 45 minutes*. So the change here between the preposition *for* and the preposition *in* is something that is engendered by the role of telicity in English syntax. Which is not a universal phenomenon; it does not apparently apply to Mandarin, does not apply to Yucatec Maya, for example, the language I've been studying in the field for a long time. But some languages do have this distinction and formal semanticists have tried to explain why that is, although they haven't had a huge amount of success at that. And by the way, Ray Jackendoff, one of the conceptual semanticists, has also come up with an insightful treatment of this problem.

Now it so happens that—well, it doesn't just so happen. Internalist, conceptual, mentalist semanticists work mostly on problems of lexical semantics, whereas formal semanticists work mostly on problems of compositional semantics, meaning the semantic impact of syntactic structure. The reason for this is that formal semantics is by definition about compositionality and on the other hand, like I said, tends to have very little to say about the issues most prominent in lexical semantics, such as semantic transfer, lexicalization, and so on and so forth, because those semantic properties can apparently only be given satisfactory explanations with reference to properties of the mind.

Now I'm going to have to, obviously I should say at this point that for many people, the opposition between these two approaches to meaning, the internalist and the externalist approach—the internalist in terms of mental representation, and the externalist in terms of representations of external reality—many people on both sides, champions of both approaches, have framed this as an issue of right or wrong, an issue of one approach being the correct one and the other people being confused and misguided. For example, Jackendoff and Lakoff have lambasted the formal semanticists as objectivists, people who supposedly necessarily believe in objective truth. Some prominent people in formal semanticists do in fact believe in objective truth and think that it necessary for semanticists to believe in objective truth and that what the mentalists are doing is basically not science.

I don't share this view at all. I'm happily working in both of these traditions, more in the conceptualist tradition because, being somebody who's interested in semantic typology, that's where most of the research has been. But I've also done some work in formal semantics. Personally I like it very much. It's just I don't believe that one excludes the other. I actually think these are complementary perspectives on linguistic meaning. The perspective of meaning in terms of reference and the perspective in terms of conceptual representation are two pictures, both of which are required for a full understanding of the phenomenon of linguistic meaning.

Now this is where we get to field semantics, leaving the preparations, the background behind. And the problem that I want to talk about centrally today is what Lisa Matthewson has called "relativist agnosticism." Lisa Matthewson is a formal semanticist who works on Native American languages, and she published a paper in the premier journal on the study of the Native American languages, the International Journal of American Linguistics. In that paper she attacks the position that she calls "relativist agnosticism," which is the view that you cannot study meaning without native speaker intuitions. You have to be a native speaker of the language in order to study meaning in that language.

This is something that I encounter a lot, and it causes me to laugh quietly. But at the same time, it also can cause a lot of frustration. I understand this view very well, I think. And I try to help people extricate themselves from this belief that it's impossible to study meaning unless you are actually a speaker of the language. But let's try to understand where this idea comes from.

To understand that, you have to take into account a little bit of the history of linguistics and more specifically the place of linguistics in academia. Which is kind of interesting, because we are sort of between the fronts. What fronts am I talking about? The fronts between the sciences and the humanities. Sciences, both the natural sciences and the social sciences are invested in observation, they're empirical approaches to the data, the truth, the reality, the knowledge.

The aim of science is to accumulate knowledge on the basis of observation, whereas the aim of the humanities is to accumulate knowledge on the basis of interpretation. And that's where the word "hermeneutic" comes in, for the method of studying the knowledge, of accumulating knowledge on the basis of interpretation. Which is the dominant method, for example, in the study of law. Because what does the study of law consist of? You look at the legal precedent or the intent of the framers of the constitution and you are trying to discern from that how a law is supposed to be applied to a particular case at hand. That's what jurisprudence is all about.

Or take philosophy—if you go take a class about Aristotle, or Marx, or Heidegger, you're going to read the words of these philosophers and you are trying to make sense of what they are saying, so you are interpreting it. That's basically your aim. Same as in art history, you are interpreting the works of Michelangelo or some other famous painter, and so on and so forth.

"Hermeneutic" is the technical term for the method of deriving knowledge from interpretation. This is a term that was coined by Aristotle in that book that I was referring to earlier, *Peri Hermeneias*, "On Interpretation." The folk etymology of this term goes back to the Greek god Hermes (in the Roman Pantheon, Mercury), the messenger of the gods. As a messenger, of course, Hermes conveys messages—the interpreter has to discern the intent of the author of the message.

So what about linguistics? Well, linguistics is situated under the social sciences. But as a matter of fact, linguistics as a modern discipline, if you take a slightly broader view and assume that linguistics got its start in the 19th century, then it did so very much as historical linguistics. In the 19th century, all of linguistics was historical linguistics, and historical linguistics was firmly seeded under the humanities.

And the move of linguistics into the social sciences is really something that didn't happen until the 1960s. You could say that the Chomskyan Turn in

linguistics, or Chomskyan Revolution, was a bit of catalyst in this development. But at the same time, the Chomskyans actually always remained in the humanities tradition. The methodological emphasis of the Chomskyan research program on native speaker intuitions as the primary source of the linguist's data, that's something very much in line with the entire humanities tradition. The people who actually moved linguistics into the social sciences, or started moving linguistics into the social sciences, were people a little bit more on the edges of mainstream linguistics, such as Bill Labov, the sociolinguist, and Roger Brown, the pioneer of psycholinguistics and child language research.

Nowadays we've had linguistics drifting more and more into the social sciences, and over the last twenty years we've seen a real empiricist turn, not just in linguistics, but in the entire cognitive sciences, which I think is a very healthy development, and this empiricist turn has accelerated the move of linguistics out of humanities into the social sciences. But a lot of our traditions are still from earlier models and that's what's responsible for this kind of thinking. Because if you assume that our primary access to the meaning of linguistic utterances is in terms of interpretation, then of course that means the best interpretation, the most accurate interpretation, has got to be the interpretation of the native speaker. So if you are not a native speaker of that language, how can you do that?

On the other hand, if you take the perspective of a social or behavioral scientist—using an utterance to convey something about something, is behavior—How do you study human behavior? You study human behavior on the basis of what can be observed and from there you try to infer the underlying knowledge of representations and so on and so forth in the mind of the speaker. There is no difference between semantics and any other form of human behavior in that respect. So from a social and behavioral science perspective, there is no difference between a semanticist studying their native language and studying any other language. Obviously, if they have no knowledge of the language, it's going to be a lot harder. But at some point they have to start from having no knowledge of the language if they ever want to get to the position where they have some knowledge.

That's true not just of the linguist, but it is also true of the native speakers themselves. Because they start out as babies, not knowing much of anything about the meanings that are expressed in their native languages. How do they get to the point where they understand the meanings conveyed in the semantic system of their language, the language they are learning? Well, they get to that point via the observation of competent speakers: their parents, their older siblings and so on and so forth. So I want to argue that the position that we as field semanticists are in is actually no different from the position that any

empirically oriented, empirically working semanticist is in, and that position in turn is not fundamentally epistemologically different from the position of a child learning the semantic system of the language in which she is being socialized.

Fundamentally, all of us, what we have to do because we "suck at" telepathy, we cannot directly share mental representations—so what we have to do is we have to reconstruct the mental representations, the senses of linguistic expressions, from the observable referential correlates, the stuff that people seem to refer to. And I'm not saying that's easy, but it is possible. So let's take a quick look at some actual examples from child language development. Basically what happens any time somebody forms a category is, you typically get this kind of U-shaped learning curve where you have your time axis, and you have the percentage of uses that's correct, and you start out somewhere around let's say 50 percent, and then the percentage of correct usage drops sharply before it goes up and finally reaches 100 percent.

Why is that? Because in the beginning all you do is you reproduce the input, you reproduce what you've heard other people do. So that means, let's say, you are going to use the word *bottle*. But you are going to use it as a proper name, you are going to use it just as the name of the particle bottle that you heard your mom refer to. Or you use the word *teddy* only for a particular teddy bear. And then, *Aha!* The kid figures out that in fact this isn't just a proper name, there's a whole category here. There's a rule, a productive rule that allows the application of this concept to an infinite set of referents. And so the kid tries to aim for the right generalization, and of course initially she aims too high, and that's why the percentage of correct usages now drops sharply.

So you get overgeneralization, such as the kid using *dog* not just for dogs, but also for lambs, cats, wolves, and cows. Why? Because she interprets this word as meaning 'quadruped.' Or—this is from Melissa Bowerman's dairies of her own daughters—*kick* being used for the kicking of a ball, but also the fluttering of a moth, cartoon turtles doing the can-can, making a ball roll by bumping it with the front wheel of a kiddy car, and so on and so forth. So in other words, at this point, what the kid is missing is that *kick* only refers to actions that are conducted with the foot. Instead, the kid assumes that *kick* refers to any kind of sharp, sudden movement.

So this process, from under-generalizations to over-generalizations to eventually the right level of generality, this is basically what Roger Brown tried to capture half a century ago with his Original Word Game, which he described as follows: "The tutor names things"—the tutor being a competent speaker of the language, the parents, the caregivers, the older siblings, any speakers the kid observes—"The tutor names things in accordance with the semantic customs

of the community. The player forms hypothesis about the categorical nature of the things named, tests the hypothesis by trying to name new things correctly. The tutor compares the player's utterances with his own anticipations of such utterances and in this way checks the accuracy of fit between his own categories and those of the player." This is fundamentally exactly what we have to do as field semanticists: we test what an expression can refer to by examining what a real or imagined situation has to be like in order for the expression to be part of a truthful description of the situation according to native speaker intuitions. I'm going to talk more about this notion of truthfulness, which is a philosophically dicey one obviously. But so fundamentally, we observe speakers use linguistic expressions to refer to the world and then manipulate the world in order to see how that affects the usages. In that way we test our hypothesis about the underlying rules in the minds of the speaker.

Here's a slightly more elaborate version of this idea: You present the speaker with a stimulus. It shows a ball and a chair, and the question is what's the spatial relation between these? So the speaker responds to this stimulus using an utterance. What's mediating between the stimulus and the utterance is the speaker's semantic system—their mental lexicon, mental grammar, and pragmatic competence. But of course these are things we as semanticists cannot directly observe, so what we have to do is we have to infer them from the relationship between the stimulus and the observable utterance. So what we do is we form a hypothesis about these mediating systems, and we test this hypothesis by changing the stimulus and finding out how this affects what the speaker is saying here.

So this means fundamentally that observational data, meaning any data that an empirical semanticist has to work from, is necessarily extensional data. It's data about possible referents in the broadest possible sense. So the fundamental goal of the empirical semanticist and the semantic typologist is to infer senses, concepts, the underlying concepts, from the observable extensions. Let me give you an example of that. So we have two sets of data collected with the Topological Relations Picture Series, affectionately known as BowPed after the two authors, Melissa Bowerman and Eric Pederson. It's a set of pictures that show different spatial configurations of objects, and the participants' task is always to identify where one particular object is located with respect to another object.

These are data sets produced by two speakers of Mexican Spanish—so both from the same variety, same dialect of Spanish. There is a vast difference in terms of this person using the preposition *en* for everything that this speaker uses *en* for, but in addition also for these kinds of attachment configurations

for which the speaker on the left uses *colgado de* or other expressions. All of these cases of support, for which the speaker on the left has *sobre*. All of this is included in the domain of the extension of *en* on this side.

Does that mean that the two speakers have different mental lexicon entries for the preposition *en*? The answer is most likely no. Rather, what's happening here is this speaker is much more specific. They are using the preposition *en* only for those instances where it's most informative. For all the other ones, where there are more informative alternatives, they use those alternatives. In other words, what's generating the narrower extension here is one of those scalar implicatures that I mentioned earlier.

So in principle, and we are going to see this in more detail later on, one of the important aspects of isolating the sense from the extensional data is to filter out pragmatic meaning components such as implicatures from the extensional data.

Over the years a number of important objections have been raised against the possibility of doing semantics on the basis of extensional observational data. So a number of different people have argued that's actually not possible. One of them was the philosopher Willard Van Orman Quine, who argued all the way back in 1960s that when you are observing an utterance, you cannot necessarily be sure what the speaker is referring to, that's the problem of referential indeterminacy. Let's say you see a rabbit running, by and the speaker goes *Blah!*, the speaker says something. So how do you know that the speaker is referring to the rabbit, the animal, rather than the fact of the rabbit running, or maybe the suddenness of the rabbit's motion, or something like that. That's the problem of referential indeterminacy.

Another important objection that has been raised—and that's important in particular in the context of Cognitive Linguistics because it has been advocated by very influential people in this field such as Lakoff and Jackendoff—that's the objectivism charge. So the claim is specifically that if you are doing semantics on the basis of referential data, you are making a commitment to the existence of an objective reality. But that doesn't make sense, because language doesn't care about objective reality—language construes reality. Therefore linguistic meaning cannot possibly depend on objective reality. That's correct, but as I'm going to argue shortly, the objection is based on a misunderstanding of the relationship between meaning and truth. It's not actually the case that meaning presupposes truth. On the contrary, the truth of linguistic utterances presupposes their meaning, depends on their meaning. So you don't actually need to be invested in the philosophical position that there is an objective reality or that we, as humans, have access to that objective reality. You

don't need to believe in objective truth in order to do externalist semantics. Notwithstanding the fact that some people who do externalist semantics actually do believe that you have to be invested in an objective truth.

But, for now, here is my main argument. Take it for what it's worth. It's not the most subtle, it's not the most sophisticated argument in the world. But my argument is, if kids can do it, so can we. Somehow, despite of whatever merits these objections might have, the kids are able to learn the semantic system of their native language. And even if you believe in innate conceptual categories, there's no question that the vast majority of meanings expressed in language cannot possibly be innate. The meanings of lexicalization expressions have got to be learned. Most of the functional category system, the meanings of function words and inflections, have got to be learned, the meanings of constructions and so on and so forth. Somehow, kids figure it out. If they can, so can we. The linguists and even the philosophers, in principle, can figure it out. That doesn't mean that I know exactly how they do it, but I think it's possible.

The last part is where I get to talk a little bit more about this controversial and very thorny issue of the relation between meaning and truth. So the idea that meaning and truth have something to do with one another is a very old idea that goes back all the way to Aristotle and probably beyond. Although people didn't turn it into a serious research program until the mid-20th century when Alfred Tarski, a very important logician, developed a formal approach to interpreting logic set-theoretically, in terms of set theory. You can take that basically as a semantics for formal languages, because logic is concerned with inference rules, and formal logic is concerned with inference rules that depend on the meanings of statements in formal languages such as predicate logic.

So what model theory does is represent the meaning of linguistic utterances—but in Tarski's sense not natural language utterances, but just formal statements, statements in some formal language—in terms of the conditions under which the statement is true. So, the basic idea is if I say, *There is a goldfish in my bathtub*, there is a set of situations in which this can be true and that set of situations is going to be distinct from the set of situations under which this other statement is true: *There is a whale in my bathtub*. Now the interesting thing about this from the perspective of natural language semanticists is that speakers have ideas about the set of conditions under which one statement or another is true. You can use these intuitions in order to work your way backwards, reconstructing the meanings of these expressions, in this case, the semantic difference between *goldfish* and *whale*.

So this is based on the Correspondence Theory of truth, which says that truth is a relation between representations and that which they represent. Namely, a representation is true if it is an adequate representation of that

which it stands for. Which is philosophically naïve. It has the intriguing property of actually predicting the intuitions of native speakers when it comes to linguistic utterances. People's intuitions for the truth conditions of a sentence are a representation, a measure, a function of their intuitions or their knowledge of the meanings of the utterance, and more specifically the meanings of the constituents, and the way these constituents are combined in the utterance. So by describing these intuitions and by testing how the intuitions change depending on the conditions under which the utterance is used. For example, the stimulus that the speaker is referring to, or the tasks that you are giving the speaker, you can make inferences about the meanings of the expressions that are involved in these utterances.

In other words, empirically, truth conditions are a way of getting at speakers' intuitions, speakers' knowledge of the meanings of linguistic expressions. Now, are they an exclusive way? No. Are they an infallible, always reliable way? No. Are they a problematic way? Yes. They are problematic, for example, because there are serious limits to the relationship between meaning and truth. For example, the truth conditions of an utterance depend on the extension of the utterance. Now there can be expressions that have identical extensions but different senses. And those expressions are going to have identical truth conditions and their semantic differences are not reflected in these truth conditions. Example: *The glass is half full* and *The glass is half empty*. Now clearly these two utterances mean different things. And yet, think about it, there is no possible situation in which one of these is true without the other also being true in the same situation. So they have exactly identical truth conditions, even though their meanings are complementary.

Similarly, *Kryten is a mechanoid* and *Kryten is a mechanoid, and he's not an anteater*. If you know anything about logic then you know there is no possible way for one of these statements to be true without the other also being true. Every possible world, every possible situation in which one of these is true is one in which the other is also true. And yet obviously, they don't mean the same thing.

To make matters much, much worse, many utterances, maybe most utterances, are neither true nor false. They in fact cannot possibly be true or false—because they are not representations of state of affairs. Of course what I'm talking about here is non-representational speech acts, in terms of John Langshaw Austin's theories of speech acts. So consider these ones. The question *Is Kryten a mechanoid?* That's neither true nor false. It's asking the addressee whether the underlying preposition is true or false. *Kryten, be a mechanoid!* (Kryten is a figure of this 1980s TV show, Red Dwarf, that I mentioned earlier.) Obviously this is neither true nor false. *Kryten, you silly mechanoid! I bet you*

five bucks Kryten is a mechanoid. So this is a performative speech act that offers a bet. And clearly you cannot continue felicitously by saying, *That's true, I agree* or *No, you are wrong about that*. None of these moves will be felicitous in this context.

In addition, there are important meaning components that are non-truth conditional, because they are pragmatic, not semantic meaning components. Presuppositions are basically propositions that speakers and addressees have to take for granted in order to make sense of linguistic utterances regardless of whether they are true or false or neither. For example, if I say *Sally chased the squirrel out of her yard*, then that implies that there was a squirrel in Sally's yard. But if I say *Sally didn't chase the squirrel out of her yard*, that also implies that there was a squirrel in her yard. If I ask the question *Did Sally chase the squirrel out of her yard?* that again presupposes that there was a squirrel in her yard. So this particular meaning component, the proposition that there was a squirrel in Sally's yard, is a presupposition, is something that we have to assume in order to make sense of the first utterance—but it is not an entailment, it's not a truth-conditional meaning component.

Similarly, implicatures—and we've already mentioned these several times today. *Floyd has watched some episodes of Red Dwarf* doesn't entail, but implicates that Floyd hasn't watched *all* episodes of Red Dwarf. In fact, if the speaker actually knows that Floyd has watched all of the episodes, then why aren't they saying so? Why are they saying *Floyd has watched some episodes of Red Dwarf*? Basically, they are probably trying to convey that Floyd didn't watch all of the episodes. And if that isn't true, then maybe they are trying to mislead the addressee, or maybe they are being confused about what's going on here. But the most likely interpretation that the speaker is trying to convey is that Floyd didn't watch all the episodes. But you only get to this interpretation on the basis of a pragmatic principle that enriches the meaning of the utterance in context, and that's a generalized conversational implicature, namely, in this case, a scalar implicature.

So we see that truth conditions are, at best, a relatively poor and problematic expression, if you will, of the meaning of linguistic utterances. Nevertheless if we assume that as empirical semanticists, we have to rely on what the speaker is referring to—on the extensional, referential content of the utterances—in order to try to reconstruct the underlying meaning, the sense, the conceptual representation in the mind of the speaker from the observed extensional uses, then the truth conditions are an important, empirical pathway that we have to rely on.

Now I've already mentioned the rejection of truth conditions as an access route to linguistic meaning by many prominent conceptual semanticists

because of the objectivism charge, because of the argument that well, you know, you cannot actually study meaning in terms of truth conditions unless you rely on some objective notion of truth. As I said before—actually this is a misunderstanding, because it's not the case that the meaning of an utterance depends on its truth. It's the other way around, the truth depends on the meaning. So when we study the meanings of linguistic utterances, we can use the speaker's intuitions of the conditions under which the utterance might be used to make a truthful statement about the world as an indirect representation of their meaning, whether or not the situation is actually true.

As a matter of fact, think about fiction, think about a page out of Tolkien's *The Fellowship of the Ring*, or I don't know, Tolstoy's *Anna Karenina*, or what have you, any sentence out of any work of fiction. You as a reader of this novel are going to have a very precise idea of the conditions under which this particular sentence in this particular novel would be true. In the story context, you are going to be able to evaluate whether the statement is true or not—even though you know it's fiction. So clearly it's not the case that in order to study meaning via truth conditions, you have to first believe that there is such a thing as truth. Nor do you necessarily have to rely on sharing an understanding of truth with the native speakers of the language that you study, although this is a little bit more iffy.

My claim is that speakers of all languages have an understanding of what you can pragmatically do with linguistic utterances, and one of the things that language evolved to do is to allow people to communicate, to exchange ideas about stuff out there in the world. People use linguistic utterances more faithfully or less faithfully, maybe misleadingly, maybe they are mistaken, and so on and so forth. At this level, everybody who uses a natural language seems to have some sort of working understanding of the pragmatics of using language. And that's something that we can study without necessarily having to answer these knotty philosophical questions about truth.

I'm claiming that the relation between meaning and truth is that the truth of a sentence depends on its meaning and the meaning is independent of the truth. Therefore it's not the case that studying linguistic meaning via the conditions that have to be fulfilled in order for a sentence to make a truthful statement requires a commitment to something like objective truth.

So in summary: Field semantics is the elicitation of semantic data from native speaker consultants and the semantic analysis of these data based on the consultants' intuitions for things such as entailments or contradictions and pragmatic felicity. This is something that we'll talk more about tomorrow. This is going to be the big topic of tomorrow. How exactly, based on what data, using what diagnostics and tests and so on and so forth, can we get from the

observed extensions to the underlying conceptual representations in the mind of the speaker?

We introduced the distinction between referential and representational content. So, on the one hand we can view linguistic meaning in terms of direct reference to an eternal state of affairs, and on the other hand we can view it in terms of external representations of the speakers' internal thoughts about the external state of affairs.

We introduced the notions of extension, intension, and sense. The extension is the set of possible referents of a linguistic expression. The sense is the underlying concept or mental representation that the speaker taps into when interpreting the utterance and that generates the set of possible referents. The intension is something that we don't need in semantic typology and in conceptual semantics, but that formal semanticists often use to have something they can formalize instead of sense, which is a conceptual notion, something that you can't formalize.

I talked about the empirical basis of field semantics. I argued that field semanticists have to infer senses from observable extensions, since they aren't mind readers, at least usually they aren't. And to achieve this they have to manipulate the real or imagined situations under which native speakers use linguistic utterances and from there uncover native speakers' intuitions about the applicability of certain expressions in reference to these particular situations. The relation between meaning and truth and the objectivism charge—truth depends on meaning not vice versa, therefore externalist approaches to meaning don't necessarily entail a commitment to objectivist notions of truth.

Once again, I'm not trying to turn you all into externalists or formal semanticists or people who believe in truth conditions. What I'm trying to say is, we're going to have to use reference, we are going to have to use extensional data as a first step in uncovering linguistic meanings in the minds of the speakers. The same way children do when they learn the semantic system of the languages they grow up to speak. That is something that is possible without philosophical commitment to something like objective truth.

LECTURE 3

Data Gathering in Linguistics: a Practical Epistemology of Elicitation Techniques

Previously, I introduced a topic of field semantics and I gave you an idea of what I think is the basis for collecting linguistic data, especially data on linguistic semantics from native speakers of languages the researcher doesn't speak in the field.

I want to elaborate a little bit more on that. The two lectures today will be about field semantics. We're going to start this morning with methods for collecting data for semantic research in the field. Then in the afternoon I'll talk about how to analyze that data, and that will also be the topic of tomorrow morning's lecture. I'm going to offer a classification of elicitation methods—methods for eliciting data in the field. To begin with, I'm going to try to clarify this notion of elicitation itself. Once we have an overview of the possible approaches to elicitation, I'm going to try to illustrate the various types with examples from my own field research.

When you're doing field semantics and in fact any semantic research on an empirical basis, meaning on the basis of observation rather than on the basis of interpretation, then that entails that you have to start from the observation of extensional data, referential data, the kind of entities, and states of affairs that the native speakers are referring to when they are using the expressions of their native language.

It shouldn't come as a surprise that I'm going to argue that any approach to data gathering in semantics, and in fact in linguistics in general, can be classified in terms of three principle components: a stimulus—that's that referential correlate that the speaker's referring to, that the speaker's talking about; a task; and a response.

Consider a sort of cartoon illustration of that idea, a very pedestrian example. This is me, asking a Maya speaker: "How do you say 'I've got to go'?" So

All original audio-recordings and other supplementary material, such as any hand-outs and powerpoint presentations for the lecture series, have been made available online and are referenced via unique DOI numbers on the website www.figshare.com. They may be accessed via this QR code and the following dynamic link: https://doi.org/10.6084/m9.figshare.11419116

© JÜRGEN BOHNEMEYER, 2021 | DOI:10.1163/9789004362628_004

the task is a translation task, and I'm giving this task to the speaker in their native language, using Yucatec: *Bix kuya'la?*' 'How is it said?' But then for the utterance that I want to translate, which is my stimulus, I'm using the contact language—that's Spanish: *Tengo que irme*. The response is a target language utterance, meaning an utterance in Yucatec: *Pa'tik im bin*, which means "I've got to go."

So that means we have these three components. My claim is that any kind of linguistic data gathering from speakers can be classified in terms of these three components. You don't necessarily have to have all of these components in place. In fact, in the simplest case of linguistic data gathering, all you have is a response—no task, no stimulus. That is the recording of spontaneous speech events, which is the kind of data that you have to rely on in Conversation Analysis in particular. Why? Because you're interested in the strategies that the speakers of the language use in order to negotiate the structure of their conversation, which is something very interesting from the sociolinguistic perspective because it reveals a lot about the underlying social structure. But if you want that kind of data, it's important that you as a researcher don't interfere with the conversation because you're going to have a certain place in society and that is going to warp—it's going to influence—the utterances that the speakers will produce. So what you have to do is extract yourself from the equation. That's why you only get a response—no task, no stimulus. There's no control whatsoever of the researcher.

Doing this in the field in an indigenous community is actually the hardest thing that you ever get to do in field research, in my experience, and I've heard that confirmed by many colleagues. Why? Because you cannot just put up a camera or a tape recorder somewhere, let's say, in a store or on a street corner and ask people to just accept that you're recording them while they are going about their business. People will find that incredibly intrusive and they will question your motives. People in industrialized nations may be more likely to accept that kind of thing because, well, we're used to researchers watching us do all sorts of stuff. But from the perspective of people in indigenous communities, that's a very strange thing to do. So this is difficult. Conceptually it's the simplest thing that you can do, but the realization is very difficult.

So the next possibility is you have still no stimulus but you do have a response to a task. This is in fact the staple, the main approach to data gathering in the tradition of fieldwork that was created, or at least is most commonly associated with, the name of Franz Boas, the founder of linguistic anthropology and in many ways also the founder of the linguistic study of indigenous languages of Americas. So the classic thing to do in the Boasian tradition is to shove a microphone in the speaker's face and ask them "Ok, tell me a story,"

"Tell me a story about the old times," "Tell me a story about how people used to live," "Tell me how this particular ceremony of yours used to be carried out." Nikolaus Himmelmann, in a seminal article on language documentation, called this "staged speech events," because they are being produced in response to a request by the researcher. Otherwise, there isn't much control of the researcher. You basically get, maybe a traditional story, maybe a historical narrative according to the linguistic practices of the community, but the fact that this happens at this particular moment in time is caused by the task given by the researcher.

Once you have all three components in place—a stimulus, a task defined with respect to that stimulus, and a response to both—that's when we can speak of elicitation. So, to make the distinction between stimulus and task a little more concrete so that people can't accuse me of being vacuous about this, here are some simple definitions. A "semantic elicitation task" is a speech act directed at the participant by the researcher intended to trigger a set of computations involving the semantic system. And the "semantic elicitation stimulus" is a verbal or nonverbal representation intended as the input of the task. Now in comprehension and judgment tasks, the input, meaning the stimulus, is an utterance, and in production tasks the stimulus is something you give to the speaker that constrains the content of the utterance to be produced. We might want to say it defines the content, but that's slightly naïve because it's the speaker who defines the content. It's the speaker who construes the stimulus in some way or another. So at best the stimulus imposes a constraint on how the speaker construes it. So then we can define elicitation itself as the collection of responses to verbal or nonverbal stimuli designed to study the respondents' linguistic competence and/or their practices of language use.

The argument I'm going to make is that we can achieve a classification of all possible elicitation techniques by starting from a classification of the possible stimulus types and a classification of the possible response types and letting the tasks simply be functions that map any possible stimulus type into a possible response type. This gives us seven possible types of elicitation techniques, starting from four stimulus types—a target language utterance, a contact language utterance, a linguistic representation, and a nonverbal representation, and the same four types for the response.

So why do you wind up with seven types rather than eight, or twenty? The reason it's not twenty is because all this stuff down here, while it's possible—you can do that; you can, for example, go from a contact language utterance to a judgment. So you can go, for example, from a contact language utterance to a judgment—but if you do that, you are no longer testing the speaker's native language competence; you are now testing the competence in the contact

language—their second language competence. So what happens here is you are going beyond target language elicitation. It may still be elicitation, but it's no longer target language elicitation. And why is it seven and not eight? Because one type occurs twice, namely, translation. Translation can lead from a contact language utterance to a target language utterance, but also vice versa, from a target language utterance to a contact language utterance. So that's why you wind up with seven types.

Now if you followed along so far, you are bound to wonder what the heck is the difference between a target language utterance and a linguistic representation. What I mean by that is that in the case of the target language utterance, the utterance itself is the stimulus, whereas in the case of a linguistic representation, the utterance, the target language utterance, is only a vehicle to convey a particular meaning that you want the speaker to do something with. I'll make that a little more concrete when we see the examples. And finally, it's important to understand that generally speaking, most of what we do doesn't just involve a single type of task, but a combination of multiple of these tasks.

Now as I said, these aren't just possible elicitation techniques for semantics. In fact they are possible elicitation techniques for *any* kind of linguistic data. All of these also find their application in semantic research. So if you are starting from a given meaning, and you want to know how this meaning is expressed in the language you are working on, you can use completion and association tasks, translation tasks, tasks of contextualized production, and description tasks. If you are conversely interested in what a given expression in the target language means, you can use judgments of entailment, or contradiction, or pragmatic felicity. You can work with explications by paraphrase or by scenario. We'll see what I mean by that. And you can use demonstrations or act out tasks.

I'm going to try to illustrate the seven possible types of elicitation techniques with examples from my own fieldwork on Yucatec Maya. But before I do that I want to give you a little bit of background on this language that I have been working on for 20 years. It's the largest member of the Yucatecan branch of the Mayan language family, which is one of the branches that split off from the common language the earliest. It's currently spoken, according to the latest census data, by around 760,000 speakers age five or older in Mexico, in the three Mexican states of Campeche, Quintana Roo, and Yucatán. Those are the states that together comprise the Yucatán Peninsula together with Belize, former British Honduras, and there are some villages in Belize where Yucatec is also spoken. I should say between the 2000 and 2005 census, the number of speakers actually dropped by about 40,000, which is scary. I'm not sure whether it reflects an actual change in the number of speakers or a change in

methodology of the census. I've never been able to ascertain this. But it is unfortunately the case that a drop in the actual number of speakers is plausible. Because what I've seen over the course of the 20 years that I've been working there is that the number of families—I work in a small village, Yaxley, which is a village of about 600 people and has been stable in the population figure for the entire 20 years that I've worked there. There's been a dramatic shift in the number of families who raise their kids in Spanish rather than in Maya, due to changes in the education system apparently.

Yucatec is a polysynthetic language; it's a purely head-marking language; it's a verb-initial language, although you don't necessarily see that in discourse, because it's a discourse-configurational language just like Mandarin, meaning topics are formally marked and it's very rarely the case that you encounter a clause with full noun phrases without one of them being topicalized. It also has a typologically highly unusual argument marking system, a split-intransitive argument marking system that's based on viewpoint aspect, something very rare in the languages of the world.

The village is right in the center of the state of Quintana Roo, about 50km from the Caribbean coast as the crow flies, although culturally the sea is really not a part of people's lives at all. These people have been tropical slash-and-burn horticulturalists for thousands and thousands of years. Their basic style of subsistence farming hasn't changed much over the millennia.

The first type of elicitation task that I want to take a look at is completion and association tasks. When I talk about completion, I mean "completion" as in completing a sentence or text with a gap, and "association" as in word association tasks. So this is the type of task where the stimulus is a target language utterance, so in this case a Yucatec utterance, and the response is also a Yucatec utterance. You give the speaker a Yucatec utterance and you ask them to complete it or to tell you something, the first thing that comes to their mind when they hear this particular expression.

The example that I'm going to give you comes from a study that my colleagues at Max Plank Institute for Psycholinguistics and I conducted about ten years ago on verbs of cutting and breaking crosslinguistically. In this particular work that I did in the field, I had a specific question that went beyond the task that we defined for the group project, which was the old Event Representation project jointly directed by Melissa Bowerman and myself. I was interested specifically in figuring out which of the Yucatec verbs of cutting and breaking were semantically specific, imposing specific selection restrictions on the theme, the object that is been being broken or cut, and which were semantically specific in terms of selection restrictions on the instrument that was being used in the breaking or cutting.

I'm going to give you a little bit of background on the general study of cutting and breaking verbs. What we did for the study is we filmed a set of 61 short video clips, which showed various scenes in which people break or cut objects. And to create this set of clips, we basically used a grid where we cross-tabulated possible kinds of objects with possible actions and/or instruments. Obviously, using a hammer or mallet to induce separation in a shirt or a piece of cloth is not the most natural way of breaking it, but we got some very interesting responses to these clips. Of all the nonverbal elicitation tasks that I've done over the years, of all the nonverbal stimuli, this is one of the ones that we got the best speaker responses to, because around the world people really like these clips. They really liked to watch these clips. They had a lot of fun because it seems to be the case that humans universally like to watch stuff getting destroyed.

People watch these clips and then we would ask a number of questions about them, such as what did the actor do in these clips, what happened to the theme—the patient, the object—in this clip, and so on and so forth. There were a number of different objectives. Asifa Majid and Melissa Bowerman were predominantly interested in the semantics of verbs of cutting and breaking and to what extent this conceptual domain is lexicalized differently in the languages of the world.

I was predominantly interested in the argument structure of verbs of cutting and breaking, meaning—well, you know, basically the extent to which the lexical semantic properties of the verbs predicted their syntactic properties across languages. There was a strong prediction out there, based on an old paper by Guerssel and colleagues that came out of the old MIT Lexicon Project in the 1980s, and I wanted to test that prediction, and it failed to confirm. Which is good, because that's an interesting finding.

So what did I do to find out which of the Yucatec verbs of cutting and breaking are semantically specific in terms of the object that's being broken and which verbs are specific in terms of the instrument that is being used? Basically, I used two different elicitation frames, or elicitation prompts. In order to get theme-specific verbs—if there are any—I asked, "I want you to tell me," so this is the question I asked for each of these verbs that people had used to describe the 61 clips. "I want you to tell me the kinds of objects that can be VERBed", and you would have to insert the particular verb that I was interested in. "If you hear that somebody VERBed something, what kind of thing are you going to think that it is that they VERBed." And to get at instrument specificity, I would ask them, "I want you to tell me the kinds of objects that one can VERB with. If you hear that somebody VERBed something, what kind of thing are you going to think it is that they VERBed it with."

Typical responses for the verb *hat*, which means 'to tear', like tear a piece of cloth or a piece of paper, basically included an object that has a fibrous structure. There is a lot of folk physics that goes into the conceptualization of this domain, which is very interesting. You get responses such as cloth, paper, leather, a plastic bag, a letter, one's hand, one's mouth or lips, shoes, and so on and so forth. This is a wide range of objects, but it's coherent. All of these are objects that you can associate with some kind of fibrous structure. As opposed to, if you ask the same question with the verb *xot*, which means 'cut,' you get, for example, rope, melons, squash, tomatoes, one's hand, one's clothes, a plank, or the table, another person, and so on and so forth, meaning this is a much less coherent class of objects. So my judgment on the basis of how much the responses would cluster around a particular conceptual type—the upshot for me was that the verb *hat* was theme-specific, whereas the verb *xot* was not.

Conversely, on the instrument-specific side, the verb *xot* would elicit responses such as, you can do it with a handsaw, a knife, a machete, a reaping hook, a hacksaw, and so on and so forth. All of these are bladed instruments, whereas in the case of the verb *hat*, you can use your hands; you can use your feet, your mouth; you can use a stick, a machete, a knife, an axe, a piece of wire, scissors, and so on and so forth. So some of these things are bladed, but by no means all of them are, and if there is a coherent theme here I can't find it. So I concluded that this verb is not instrument-specific.

The original literature on this type of approach comes from language production research, of course, and research on lexical processing. More recently there's a great paper by Nick Evans and David Wilkins on semantic metaphors of perception verbs in Australia languages where they used word association techniques.

Translation leads from a contact language utterance to a target language utterance, or vice versa, from a target language utterance to a contact language utterance. There are two fundamental problems with translation tasks. One is you have insufficient control over how the speaker construes the stimulus. This is a general problem with any kind of elicitation. Specifically in the case of translation, a compounding factor is that by definition, since it's a contact language that is used in translations tasks, it's not the speaker's native language, so they likely have their own nonnative-speaker variety of this language. It may not be the researchers' native language either, so that means the researcher also uses some second-language-speaker variety of this language, so basically what's going on here may be quite fuzzy.

An additional problem that has long been noticed in connection with translation tasks is that the speaker may attempt to translate not just the meaning of the stimulus utterance, but also the form, the phenomenon known as

"interference." So they may try to produce a target language form that corresponds to the form of the stimulus utterance.

So the example I'm going to use is Östen Dahl's Tense-Mood-Aspect Questionnaire, which is not just an example of translation questionnaires themselves, but also an example of a very valid attempt at solving the problems inherent to translation tasks. Now I don't know how familiar people are with Östen Dahl's typology of tense-mood-aspect systems and his questionnaires. It's a questionnaire consisting of originally around 156 scenarios that Östen Dahl and his collaborators and students used to collect data on tense-mood-aspect systems from a very large number of languages. It's a classic study and a very useful tool, I've found. As I mentioned yesterday, I did my dissertation research on the tense-mood-aspect system of Yucatec, so for me this was a very valuable tool. I also use it in every field methods class I teach, because it gives you valuable data, not just on the tense-mood-aspect system of the language, but also on the basic verb grammar, which is pretty much the core of the grammatical structure of every language.

What Dahl does in order to overcome those two fundamental problems of translation-based elicitation is he accompanies every stimulus utterance with a scenario in which this utterance is supposed to be interpreted, a scenario that's supposed to give context to the stimulus utterance and thereby clarifies the intended meaning.

So for example, say we want to ask the speaker for the translation of "He will be writing letters." So you're getting imperfective or progressive aspect, future time reference, and an eventuality type that is unbounded, atelic, an activity, writing letters rather than one particular letter.

In order to clarify this to the speaker you're using a context that defines the reference time of this utterance. And the context is "What your brother DO when we arrive, do you think?" Now this is actually not a grammatical utterance in English, and I will explain in a second why this is the case. Basically the context introduces a scenario in which we are talking about something that will be happening in the future. So we get a future reference time. It will be happening at the moment that we arrive. So the question is what activity will the brother be engaged in at the moment when we arrive? So this gives you imperfective or progressive aspect.

So the reference time, which is crucial in research on tense-mood-aspect systems, is controlled by the context, by the scenario. Generally speaking, the best approach to running this task is to let the speaker translate, not just the stimulus utterance, but also the context, the scenario. Just have them translate the whole thing. It's the most natural way to do that.

Now, why is it that the context and the target, the stimulus utterance, are given in these ungrammatical forms using these infinitives in capital letters? That's Dahl's attempt at preventing the speaker from translating the form of the contact language stimulus utterance. So instead of giving them finite verb forms that may influence the finite verb forms the speaker produces in response, Dahl uses infinitives. And the idea is that the speaker tries to figure out what it is that the researcher is trying to get at and then inserts in their mind the appropriate forms in the target language.

Now the first solution, meaning the use of these context scenarios, was very successful. It's a great idea; it's beautiful; it's contributed a lot to the success of this tool. The second solution, the use of these infinitives, is not so great. Because what it triggers almost universally for almost anybody who has ever tries it is a lot of confusion, especially when you are working with illiterate speakers. So you have to read this out to them, which means you're giving them an ungrammatical sentence in the contact language—"What your brother DO when we arrive, do you think?"

Now what's the speaker supposed to do with that? How are they going to interpret that ungrammatical sentence? So that's a problem, and I can't say that I've ever been able to solve this problem. It's easy to go around it if you are working with speakers who are literate. You can show them a printout, or you can show them the items on your computer screen, and they understand okay, there is something funny going on with these infinitives here because they appear in capital letters. If you are working with illiterate speakers, it's going to be tough.

This is the use of an utterance in the target language or in a contact language that serves, not itself as a stimulus, but rather as a vehicle for a particular meaning that you are trying to get across. In this case that's exactly what happens in the Dahl Questionnaire. So the context is not actually an utterance, the translation of which you are interested in. Rather it's something that you are using to set up a scenario that serves as a context in which the speaker is supposed to produce the target utterance.

So that leads us to Type 4, which is the description of nonverbal stimuli. And this is really the main stay in my narrow field of specialization, semantic typology. Most of the research in semantic typology has been based on descriptions of nonverbal stimuli. For example, take Berlin and Kay's seminal study of basic color terms: those Munsell color chips that they used to elicit the semantics of color terms across languages are a classic example of nonverbal stimuli.

There's a wide range of different types of nonverbal stimuli that people have been using. We already saw examples yesterday of the BowPed line drawings,

which show different spatial configurations of objects. We saw one example of a task that involves photographs, the Ball and Chair task.

You can use comics and picture books. A very famous example is the children's picture book "Frog Where Are You?" that Ruth Berman and Dan Slobin have been using in cross linguistic research on child language. There are also video clips such as the Cut and Break clips, which we already addressed.

BowPed consists of 71 line drawings, featuring so called "topological" relations. This is "topological", not in the mathematical sense, but in the sense that was introduced in child psychology, developmental psychology, by Jean Piaget and Bärbel Inhelder, and that was in the 1950s, in a famous book called "The Child's Conception of Space." So they had the idea that the simplest kind of spatial relations are topological relations, meaning relations that are perspective-free, that don't depend on a particular viewpoint, such as support, inclusion, and so on and so forth.

So there are a number of problems that are associated with the use of nonverbal stimuli. First of all, it's very important to be specific about the task you are giving the speaker. Nonverbal stimuli are kind of deceptive in that way—deceptive for the researcher, that is. When I see students work for the first time with nonverbal stimuli, a lot of them do the same thing, which is they assume "Ok, great, I'm just going to show this video clip, or I'm going to show this picture, and then the speaker will produce what I want to hear, and I'm done."

And of course that doesn't work. Because what happens is, let's say in the case of the BowPed pictures, your goal is locative descriptions. So let's say in each picture one object is selected, it's marked by an arrow—that's supposed to be the figure of a locative description you are trying to elicit. The other object is the ground with respect to which you want the speaker to locate the object. So therefore you ask the speaker, "Ok, tell me where the ___ is." And the speaker looks at you, and looks some more, like you are not quite with it maybe, and the speaker says "It's right here in the picture. Can't you see?" So what happened here of course is that the speaker was absolutely right. The mistake was entirely on the part of the researcher. The researcher gave a very imprecise instruction to the speaker.

So here is a more appropriate version, although this is admittedly very elaborate. This is another example of an elicitation prompt or an elicitation frame. Imagine this is something that you're telling the speaker before they produce the utterance you want to hear. Imagine you are talking to somebody who is looking for the figure, so in this case imagine you are talking to someone who is looking for the cup. This person knows where the table is, but doesn't know where the cup is. You know where the cup is, but neither of you can actually see

the cup and the table right now, and the person asks you "Where is the cup?" Imagine you want to tell that person where the cup is. How do you respond?

That's very involved, and especially if you have to repeat it every time with every new picture, it's going to be very tedious, and it's going to get the informant, the consultant annoyed and exhausted very quickly. So what's important is that you make sure that the speaker gets the hang of this, understands what you are trying to do, and once that is the case, you don't need to repeat it over and over again. You just once in a while might want to make sure that the speaker still remembers.

Next, visual stimuli have been used not just in production tasks but also in comprehension tasks, especially in child language research. In fact a lot of this work with nonverbal stimuli was pioneered in child language research and from there has been exported to cross linguistic research on indigenous languages. So you use nonverbal stimuli for example in verification tasks where you ask a speaker "OK, here is an utterance. Is the scenario that is shown in this particular picture here a good example of what this utterance says?" "Is this utterance true?"—that's why it's called verification task. "Is this utterance true in this particular picture?" And you can do that even with small children. You can for example use a hand puppet of Cookie Monster or something like that, and you show them a picture and you say "Well, OK, Cookie Monster says that there are two cookies in this picture. Does Cookie Monster tell the truth or is Cookie Monster confused again?"

Nonverbal stimuli are used in matching tasks, which are often in turn used in referential communication tasks, tasks that involve two speakers of the language interacting with one another. And we'll see examples of that as we go along. In fact here [pointing to diagram on slide] you see this is the basic outline of a picture-matching task, such as the Ball and Chair task, which I did mention before, which is example of a referential communication task.

So in this case you have two speakers and a screen between them. Actually, let me go back a little to the pictures, the photographs that I showed you from my field site. This photo shows you two speakers doing the Ball and Chair task, and you see I'm using my suitcase, the suitcase I use to haul my field equipment, the camera and so on and so forth, as the screen. And each of the two speakers has a set of 12 pictures, 12 photographs in front of them, each photograph showing a ball and a chair. They only differ from one another in the location of the ball with respect to the chair and the orientation of the chair. And one speaker picks a picture, describes it, and the other speaker has to find the corresponding picture in their set, and the order of the pictures in the two sets is different.

FIGURE 3.1 Layout of the Ball & Chair referential communication task

The point of the referential communication task type is to force two speakers to be referentially maximally explicit. So on the one hand, because you have an interaction between two speakers of the language, they themselves get to choose what is the most appropriate strategy to solve this task in their native language. So in particular, if you are interested in preferences in communicative behavior, this is a great task. So in the line of research on spatial frames of reference, we are interested in what types of reference frames speakers of different languages prefer to use, so therefore this is a very productive type of approach.

And on the other hand, the screen, the fact that these two speakers can't share visual attention, forces them to be linguistically maximally explicit about the referential content of their utterances. Which of course renders this type of task very artificial, because generally speaking this is not a natural communication context. Generally speaking, when people interact, they share a visual field. They don't necessarily talk face to face. That's a common mistake that European researchers in particular have been making, since in the Euro-American cultural context, face to face is the default of dyadic conversation. That's not the case in many other cultures. But still, you know, generally speaking, interlocutors are in the same place, they share a visual field, and so this is not a particularly natural type of task.

What does that mean in terms of ecological validity of the data that you are collecting with this task? Of course what it means is you never want to rely, when analyzing the linguistic practices of this community, on the results of such an artificial task alone. You always want to check whatever you get out of that task against other more natural, more ecologically valid observations.

Finally, the interpretation is very important. This is the general point that I raised before: whenever you are doing elicitation, you have to be aware that ultimately the speaker is not responding to the stimulus, or rather that their response does not directly reflect the stimulus *itself*, but rather their conceptual *construal*, their interpretation, of the stimulus. That interpretation may be subjective to cultural conventions.

The classic example that I like to use to illustrate this is an anecdote that comes from my colleague and friend David Wilkins. David did a study on the acquisition of motion verbs in children learning Arrernte, a Pama–Nyungan language of central Australia, of the area around Alice Springs. He created a series of line drawings that were supposed to illustrate different manners of motion, such as a galloping horse. The Arrernte children looked at this picture and they got very sad, and they said, "Well, that's a dead horse."

So that took the researcher a bit to figure out. It turns out of course that the default perspective for visual representations in many aboriginal cultures in Australia is not horizontal as in western visual representations. I believe that's also the case in China. But rather, it's a top-down, birds' eyes perspective. So if you imagine that you're seeing this horse from above, that means the horse has got to be lying on the ground. If you know something about horses, horses generally don't lie like that unless they are not that well. So then what about all these dust clouds here, which David added to the picture in order to animate the horse? That was supposed to show that the horse was running very fast. Well, the kids interpreted that as the traces of rigor mortis that the horse had left behind in the dirt when it died.

Of course cultural conventions for the interpretation of visual representations don't just change from place to place—they also change over time. So even in European culture, conventions for visual representation have changed a great deal since the Middle Ages, and a great example for that is the Bayeux Tapestry, which shows the events of the Norman Conquest of England. Such as in this case, in this sequence here, the death of Edward the Confessor.

And basically the way this tapestry is supposed to be interpreted is sort of as a cartoon strip. Except that the different panels of the cartoon are not separated from one another. So in effect what's happening here is that temporal sequence, temporal proximity is translated as spatial proximity, so all the events of the Norman Conquest are basically shown in a spatial sequence. So it looks

as though these things happen all at the same time in different but adjacent places, that's the way we would interpret that nowadays, but the way it was intended is to show a sequence of events.

The next type I want to look at is the elicitation of judgments. So in this case the stimulus is a target language utterance, and the response is a judgment about this utterance. Which can be a judgment of syntactic well-formedness; it can be a judgment of semantic interpretability; a judgment of contradiction versus the absence of contradiction; it can be a judgment of pragmatic felicity, and so on.

The example I want to give is one of testing event descriptions for telicity. So what is "telicity"? It's the property of event descriptions of having an inherent goal that has to be achieved in order for the event being described to be realized. And some event descriptions have this property and others don't. So we already saw the example of *Sally watched an episode of Red Dwarf* versus *Sally watched Red Dwarf*. You can be watching Red Dwarf without there being any inherent endpoint that you need to reach in order to have accomplished something; as soon as you start watching Red Dwarf, you have watched Red Dwarf. In contrast, if you are watching an episode of Red Dwarf, then until you've actually finished that episode, you haven't actually watched the episode. So that would be an example of a telic description.

Event descriptions can be classified as states, activities, accomplishments, and achievements, which goes back to Zeno Vendler. Telic descriptions are the accomplishments and the achievements. The states and activities are atelic. The other two properties that define this four-way classification are the properties of dynamicity and durativity, the distinction between durative and instantaneous events.

I wanted to test Yucatec event descriptions for telicity. Now, for English, people have observed over the years a number of correlates of telicity. This table goes back to a book by David Dowty from 1979. And it points out, for example, that telic descriptions use *for*-type adverbials, whereas atelic descriptions use *in*-type adverbials. Telic descriptions go with *finish* whereas atelic descriptions don't. Both of them go with *stop*. But if you apply *stop* to a telic description, then the implication is that the eventuality was never realized.

Now, these particular tests that I just mentioned actually rely on *syntactic* properties of event descriptions in English. It turns out these syntactic properties are not universal. For example, I believe that the distinction between *for*-type and *in*-type adverbials does not apply to Mandarin. The syntactic distinction also doesn't apply in Yucatec. In fact, Yucatec has *no* syntactic correlates of telicity, so there's likewise no distinction—well, there *is* a distinction

TABLE 1 Diagnostics of the Vendler classes for English (Dowty 1979: 60)

Criterion	States	Activities	Accomplishments	Achievements
meets non-stative tests	no	yes	yes	?
has habitual interpretations in simple present tense	no	yes	yes	yes
φ for an hour, spend an hour φing	OK	OK	OK	bad
φ in an hour, take an hour to φ	bad	bad	OK	OK
φ for an hour	yes	yes	no	N/A
∴ φ at all times in the hour				
x is φing* entails *x has φed	N/A	yes	no	N/A
complement of *stop*	OK	OK	OK	bad
complement of *finish*	bad	bad	OK	bad
ambiguity with *almost*	no	no	yes	no
x φed in an hour	N/A	N/A	yes	no
∴ x was φing during that hour				
occurs with *studiously, attentively, carefully*, etc.	bad	OK	OK	bad

between *stop*-type aspectual verbs and *finish*-type aspectual verbs, but all Yucatec event descriptions are in fact compatible with the latter.

So what I did instead is rely on entailment pattern that we already looked at yesterday. So here is another example of the same pattern: if it's true that Floyd was pushing the cart, then that means it also must be true already that he pushed the cart. As soon as it's true that Floyd was pushing a cart, it's also true that he pushed the cart. So that's an entailment. That makes "push a cart" an atelic description.

In contrast, if it's true that Floyd was drawing a circle, it does not necessarily follow that Floyd drew a circle. That depends on whether he actually completed the circle. So "draw a circle" is one of these eventualities that have an inherent endpoint. In other words description is telic.

So this is what I used with Yucatec speakers. I set up a scenario in which an eventuality of the requisite kind was interrupted, and I asked, "At this point,

can you say that the eventuality has already been realized?" Now, this is of course a very abstract thing to do. It requires a lot of focus on the part of the speaker, and in addition you have to find a scenario in which this makes sense for the speaker. The only way to do that is to negotiate a possible scenario for each verb that you want to test with the speaker. Negotiate with them an example, a context, or scenario in which an action or event of the particular kind can be interrupted.

So here [pointing at screen] is an example. In this case the verb I want to test is the verb *k'áay*, which means 'sing.' The context that we came up with this particular speaker is shown in (6):

(6) *Pedro=e'* *táan* *u=k'àay,*
Pedro=TOP PROG A3=sing\ATP
'Pedro, he was singing,'
káa=t-u=k'at-ah *u=báah* *Pablo.*
CON=PRV-A3=cross-CMP(B3SG) A3=self Pablo
'(when/and then) Pablo interfered.'
Pedro=e' *t-u=p'at-ah* *u=k'àay.*
Pedro=TOP PRV-A3=leave-CMP(B3SG) A3=sing\ATP
'Pedro', he stopped singing.'
Be'òora=a' *ts'o'k=wáah* *u=k'àay* *Pedro?*
now=D2 TERM=ALT A3=sing\ATP Pedro
'Now, has Pedro sung?'

The answer from an English perspective that would be expected here is the answer 'yes,' because in English *sing* is atelic, it's an activity, it doesn't have an inherent endpoint. But on closer inspection, of course, *sing* has a cognate object, *song*, and even for an English speaker, you may assume that when you sing, you generally sing a song. A song has an inherent endpoint, so you haven't actually sung until you completed that song. It's just that for English speakers, this interpretation is not very prominent. For Yucatec speakers, this is much more prominent. Why? Because Yucatec, like many Mayan languages, does not have a lot of basic "unergative" verbs, basic activity verbs. Rather, activities, in Vendler's sense, are lexicalized primarily as nouns, or alternatively, one could argue that they are lexicalized in a class of parts of speech that can be used as nouns or verbs without overt morphological derivation either way. "Action nominal," they're called, among Mayanists.

So this morpheme *k'áay* does not just mean 'sing.' It also means 'song.' And this bound pronominal marker here [i.e., *u=* in *u=k'áay*] can be interpreted not just as an ergative marker of the single argument of this verb, but also as a

possessor marker. Which is the reason why it is called not "ergative," but rather "A3." It's a "Set-A marker," as Mayanists call it. It conflates the function of ergative and possessor. So that means the first line of (6) does not just mean 'He is singing,' but also 'His song is going on,' and the question at the end of (6) does not just mean 'Has he sung?' but also 'Did his song finish?' Because what I'm translating here as a perfect aspect, *ts'o'k*, the so-called "terminative," is actually a grammaticalization of a verb that means 'to end.'

Because of this ambiguity, the interpretation 'Has Pedro sung?' and 'Did Pedro's song finish?' are both salient to the Yucatec speaker. As a result a lot of speakers will say 'No' in response to this question. Which makes another very important point about elicitation in general, which is, when you are interpreting the speaker's response, you cannot take it at face value. The speaker's response becomes a data point of your analysis only once you have taken into account how the speaker interpreted the stimulus, and how the speaker interpreted your task, and how the speaker intended their own response to be interpreted by you.

Now when you are doing this complicated thing of testing for possible scenarios which instantiate and satisfy the semantics of a target language expression, it's usually a big help if you can use nonverbal stimuli to help clarify that scenario. So in that case you would use a combination of a linguistic utterance and a nonverbal stimulus in order to get the best possible and most accurate judgment by the native speakers.

I'm going to illustrate that with an example of research I've done on "verbs of inherently directed motion," as Beth Levin calls them, or "path verbs," as Len Talmy calls them, meaning verbs such as *enter* and *exit* and so on. And the background here is an important study that my former colleague Sotaro Kita did on Japanese, where he showed that Japanese *hairu* and *deru*, 'enter' and 'exit,' respectively, actually are more appropriately translated into English as 'become inside' and 'become outside'—because they don't necessarily entail that the figure, the entity whose motion is at issue, did in fact move, rather than the ground.

So let's say you start out with a square inside a circle and you wind up with a situation where the square is outside the circle. One possibility is of course that the square has moved, and in that case you can use *deru*. But another possibility is that the circle has moved and now the square came to be outside of the circle because the circle moved. And in that case, too, you can use *deru* in Japanese—but you can't use *exit* in English.

So I wanted to know of which kind the apparent path verbs of Yucatec were, whether they were of the Japanese kind or of the English kind. Whether the Yucatec path verbs, or location change verbs, were of the kind the English path

verbs are or whether they were of the kind the, well, at least *hairu* and *deru* and possibly other location change verbs of Japanese are.

So I used a series of video clips that was designed by Steve Levinson, the so-called "Motion Verbs Clips." So in this case, can you say that the ball entered the enclosure? [Audience says no.] But Yucatec speakers will say just that. And in this case [plays clip of a wooden ramp sliding under a ball], can you say that the ball ascended the slope or rolled up the slope? And again English speakers will tend to say no, but we'll see in a moment what Yucatec speakers do. So when you present this clip to a Yucatec speaker and you ask them 'Is it possible to say (7)?' you are using the clip to define the situation and the utterance to ask the speaker 'Is this a possible accurate truthful description of this particular scenario?' So it's a combination of two stimuli, the utterance and the video clip, basically what we saw before the break as a verification task. It's a kind of verification task.

(7) *H-na'k* *le=chan* *kanìika*
PRV-ascend(B3SG) DET=DIM marble
y=óok'ol *le=tàabla=o'*
A3=top DET=plank=D2
'The little marble ascended to the top of the plank'

Out of context, Yucatec speakers will tell you 'No, that's not true.' But that's not the whole story, as we are going to see in a moment. So the response to this typically is going to be negative—'No, it did not ascend to the top of the plank.' But the reason for this is that this utterance out of context will trigger a stereotype implicature, meaning a pragmatic inference that enriches the interpretation of the utterance, even though it was not actually part of the utterance's semantic meaning, and that implicature will interpret the utterance to the effect that it was in fact the ball that moved, which isn't the case, it's not true in this video clip. Therefore the speaker will say no. So in other words, the point is quite simply that Yucatec speakers, just as English speakers or probably Mandarin speakers, will assume that it is much more likely when they hear an utterance like this that it was the ball that moved rather than that it was the plank that moved.

But that is not in fact part of the semantics of the utterance. It's just an implicature. So that implicature can be defeated, and one way to defeat it is to simply deny right up front that it was the ball that moved, or rather assert right up front that it was in fact the plank that moved. So you can say (8):

(8) *Le=chan* *tàabla=o'* *h=péek-nahih,*
DET=DIM plank=D2 PRV-move-CMP(B3SG)
káa=*h-na'k* *le=chan* *kanìika*
CON=PRV-ascend(B3SG) DET=DIM marble
y=éetel *che'* *te'l* *y=óokol=o'*
A3=with wood there A3=top=D2
'The little plank, it moved, and the little ball and the stick ascended to its top'

This is something that Yucatec speakers overwhelmingly will find perfectly acceptable as a description of this particular video clip. And notice, of course, if it was the case that motion of the figure was actually part of the entailments, the contribution that this verb makes to the overall meaning of a sentence in which it occurs, then (8) ought to be a contradiction—and Yucatec speakers will not say that this is a contradiction. They will find this perfectly acceptable. Notice first of all—again, (8) is something that I didn't simply make up and give to the speakers. It's something that I had to negotiate with each speaker. So if you are doing this kind of task with five, or six, or seven speakers, you have to negotiate an appropriate utterance with every speaker.

So this can be quite involved and time-consuming, and furthermore it instantiates another type of task, our type number six, which leads from an utterance as stimulus to a verbal representation. So in this case, what do I mean by "verbal representation" again? The verbal representation itself is an utterance, but it's not the form of the utterance, syntactic or morphological form, or even the words that occur in it, that are part of the response, it's just the meaning that is the part of the response.

In other words what I'm doing here is, I'm asking the speaker to describe to me a scenario in which something could be said, a particular utterance could be said. And this is how I got this as a possible description that's consistent with the situation shown in the video clip. I got this by negotiating with the speaker "can you think of a case, a scenario, a situation in which somebody might truthfully say 'the little marble ascended to the top of the plank'" with respect to this video clip, talking about the scene that's shown in this video clip. And then the speaker says "well, yeah, you know, we can do that. You just have to say actually the plank moved and then the video clip—the ball, the marble ended up in top of the plank."

This is an approach that I've used a lot, meaning asking speakers "Can you think of a scenario? Can you describe to me a scenario under which somebody might say this kind of thing, that it would be a truthful statement about

this particular situation here?" And that's something you cannot do with any speaker. Some people have a knack for this kind of thing. It requires a little bit of imagination. It requires a little bit of verbal intelligence in order to be able to do that, and some people have a knack for this kind of thing.

It doesn't require any education. People don't have to have education and they don't have to have training. Some people just have a knack for this kind of thing and some people don't. Other people, other consultants are very good at producing paradigms of inflectional forms for you; some people are very good at transcription, and so on and so forth. There's a range of different skills that make native speaker consultants, good consultants. This is one skill that I as a semanticist value very, very highly.

And finally, the last type is demonstrations and act-out tasks. So that leads from a target language utterance to a nonverbal representation that instantiates the target language utterance. Mayan languages have a very rich set of roots, not necessarily verb roots, which lexicalize what Mayanists typically call "positions," but which are more appropriately called "dispositions" because, while the prototype of these kinds of configurations maybe positions, postures of animate beings, humans and animals, the majority of the roots actually lexicalize spatial properties of inanimate objects. And they are not locations. They are rather what you might want to call "manners of location." That's a metaphor that a former student of mine came up with. So these roots lexicalize properties such as support, suspension, blockage of motion, orientation (mainly in the gravitational field), shape configuration of parts of the figure with respect to one another, and so on and so forth.

So, for example, what that means in practice is Yucatec has seven or eight different roots—all of which mean 'hang,' and they describe different kinds of hanging configurations; seven or eight different roots that mean 'lean'; seven or eight different of roots that mean 'be stuck on or in,' and so on and so forth. If you think that's a lot, Yucatec has altogether maybe 160 or 170 of these roots. Other Mayan languages have 700–800 of them.

So this is a very rich semantic field in Mayan languages, and of course most of these concepts are not lexicalized in better-studied languages. They are not lexicalized at anywhere near this level of specificity in languages such as English or German or Spanish.

So in other words what I'm saying is when you are trying to describe something like that, where a whole class of lexical items, a whole category doesn't have counterparts in languages that you are familiar with, you don't even know where to look for the possible meanings of these items. You don't know the dimensions of the semantic space. You don't know the dimensions of contrast that you need to assume in order to describe the meanings of these lexical items.

So what I wanted to do is get a handle at this. What I used was an approach that was pioneered by Brent Berlin in the 1960s in his famous research on numeral classifiers in another Mayan language, Tzeltal. It's a two-pronged approach, or a two-phase approach. During the first phase, I produced all the roots that I knew with a series of, with a number of five or six speakers, and I asked them each time, for each root, I asked each speaker 'Tell me all the different objects that you can think of that can be said to be in this particular disposition,' the disposition described by this particular root.

Then I compiled the responses, and I sorted them according to the classes of objects and I came up with the 20 most frequent object classes. Then I got objects that would instantiate the 20 different classes—props for the elicitation. And I arrayed all roots that occurred with a particular class of objects in a group, and I asked each speaker to show me contrastively all the different dispositions the object has to be in so that it can be said that a particular root applies to it.

What I'm doing is I'm giving the speaker all the different roots that, according to the first phase, can occur with this type of object with rope, and I'm asking them 'Ok, so, now show me how the rope needs to be manipulated, what configuration the rope needs to be in such that it can be said that it's in this particular disposition, described by this particular root?' So in other words, the response is what the speaker does with the rope. So that's what we call a "demonstration." The speaker demonstrates the referential content of the root in this particular situation.

An act-out task is something—the situation is the same thing: the speaker demonstrates the referential content; but, in this case, they do it with their own body. So let's say you're studying manners of motion, and you ask 'Ok, so this particular thing, can you show me what it means to move in this manner?' And the speaker begins running in place. That would be an example of an act-out task. If you're studying cutting and breaking, and you want to know the particular manner of action that associates in this particular verb, you ask the speaker and they make a smashing gesture—that would be an example of an act-out response.

I have two points in the discussion that I want to make. One is the difference between elicitation and experimentation—because there's an unfortunate tendency to call anything that involves nonverbal stimuli "experimental research." We don't need to talk about where this comes from, but fundamentally it's based on a misunderstanding or a loose understanding of what it means to do an experiment.

In a narrow sense, an experiment always involves a hypothesis that you are testing. Elicitation does not, at least not necessarily. You could say that as soon as you use elicitation to test a hypothesis, it becomes experimental research,

but elicitation isn't inherently experimental research, whether you use nonverbal stimuli or not. It doesn't matter. It's irrelevant.

And very important, when you are interpreting elicitation responses, you have to take into account how the speaker construed the task, how they construed the stimulus, and how they intended their response. We've seen examples of all of this. I call this the "Golden Rule of Elicitation": An elicitation response only becomes a data point in the reconstruction of a speaker's linguistic competence once the speaker's interpretation of the task and stimulus and the intended interpretation of response have been ascertained.

For a very quick summary, linguistic data collection techniques can be classified in terms of three components: a stimulus, a task and a response. Methods for eliciting expressions of a given meaning include completion and association tasks, translation, contextualized production, and description tasks. Methods for eliciting meanings of a given expression include the collection of judgments of entailment, contradiction, and felicity, the explication of the meaning by paraphrase, or by scenario, and demonstration and act-out tasks.

The epistemology of elicitation, well that's something we actually didn't talk about. Fundamentally, native speakers apply their linguistic knowledge to solving a certain problem, and the researcher reconstructs that knowledge based on the observation of the solution. This is something that I think is pretty much unique to linguistics and cultural anthropology.

Fundamentally the way I like to think of this is, imagine you are a Martian scientist and you want to study the knowledge of human car mechanics. The way you do that is, you buy a number of cars, and you muck them up, and you give them to these mechanics, and you study what they do to fix them. That's basically what we do as linguists in the field. We break stuff for the native speaker, and we get them to fix it, and from there we try to infer their native speaker knowledge.

And finally the "Golden Rule of Elicitation": A response become a data point in the reconstruction of speaker's linguistic competence once the speaker's interpretation of the task, the stimulus, and the response have been ascertained.

LECTURE 4

Sources of Evidence: Semantic and Pragmatic Diagnostics

As best as I recall, I was going to make an argument to the effect that the data that researchers collect using the kind of methods that we talked about this morning is extensional data, meaning it's data that illustrates usages of linguistic expressions in reference to a particular entity or state of affairs. But at best this data gives us information about the referential content of the expressions that we've been recording or otherwise collecting. But it doesn't yet give us information about the underlying meaning in the sense of the conceptual representations in the mind of the speaker that these expressions map into and that interpret them cognitively.

And the argument I wanted to make is that in order to get there, in order to get at the actual semantics, the sense, rather than the mere extension, we need more than positive evidence, we need negative evidence. So I was going to start out with the distinction between positive and negative evidence, and that has to do with the fact that the goal of linguistic description is to produce a record of the competence of the speakers, the procedural knowledge that allows a competent speaker to be just that, a competent speaker of the language.

And that means we have to have some representation, not just of what the speaker actually does in the contexts in which we study them, or what the speaker did when we observed them—for example, during the elicitation session, such as those we talked about this morning—but we actually need a sense of what the speaker can do, and what the speaker cannot do, or what the speaker won't do, and that requires negative evidence.

So in a sense what we're engaged in when we're trying to study language in the field—when we're trying to produce a record of the language, a language description—is a transformation of the observable procedural knowledge of the speakers, or rather the products of that procedural knowledge, into

All original audio-recordings and other supplementary material, such as any hand-outs and powerpoint presentations for the lecture series, have been made available online and are referenced via unique DOI numbers on the website www.figshare.com. They may be accessed via this QR code and the following dynamic link: https://doi.org/10.6084/m9.figshare.11419119

© JÜRGEN BOHNEMEYER, 2021 | DOI:10.1163/9789004362628_005

declarative knowledge, declarative knowledge of us as scientists, as linguists, as researchers who study these languages. I believe uncovering the underlying knowledge of the speaker, even if that knowledge itself is procedural, and if we want to make declarative statements about it, we'll need negative evidence. The reason I'm emphasizing the role of declarative knowledge is the fact that the language description itself is declarative knowledge.

Let me try to be a little more explicit about this. So you know that any competent speaker of any language is going to be able to use the grammar of that language without actually being able to explain to you why they do what they are doing. Otherwise we would all be unemployed; there would be no job for us as linguists, because if the native speakers could explain themselves what they are doing, then you know, that's it. They could just produce the description themselves.

The difference between what the speaker actually does and our record, our account of the underlying cognitive processes—that's precisely the difference between procedural and declarative knowledge. Because the knowledge the speaker has of the grammar and lexicon and the practices of use of that language is just procedural knowledge. It's knowledge of how to do things, not knowledge of facts that you can state verbally, such as verbal generalizations, the kind of generalizations that language descriptions consist of.

That's why I argue we need negative evidence. That constitutes an important difference between language description and language learning by children. Earlier, I argued that fundamentally as field researchers studying the semantic system of the language, we're not in a different position from that of the child learning the semantic system of their native language. Except that there is a difference when it comes to describing the semantic system, because that's something the child doesn't have to do.

So the child presumably is going to be able to learn all the procedural knowledge that's required to speak the language on the basis of positive evidence alone, whereas the researcher is not going to be able to describe the language on the basis of positive evidence alone. There's been a long-standing debate going on for many decades in language acquisition research, or child language research, on the role of negative evidence. Generally, most people in the field assume that children don't have a whole lot of negative evidence available and obviously they're still able to learn the language they are growing up with just fine.

What do I mean by negative evidence in this case? Basically suppose the child makes a wrong generalization. Remember that U-shape learning curve that we talked about yesterday. The child has figured out that there is a productive category here, but hasn't quite zeroed in on the appropriate rule. The

category the child is hypothesizing is underlying the adult's uses is actually too broad, so the child overgeneralizes. The problem is that it's not typically the case that every time the child does it, there's some competent speaker at hand who will correct the child. That's what I mean by negative evidence in this case. So, you know, the child uses the word *dog*, not just for dogs, but also for lambs and cows and cats and what have you. It's not the case that every time that happens, there is an adult around who is going to tell them, *No, actually that's not a dog, that's a cow: big animal, gives milk, moos; dog: small animal, barks, bites, furry, etc. etc.*

So somehow oftentimes the child does have negative evidence, but apparently, she's going to be able to get by, even without a whole lot of negative evidence. How exactly that works is a big mystery, and a lot of ink has been spilled and a lot of research has been devoted to this question, but it's not the case that a definitive answer has been found.

However, this constitutes a fundamental difference between semantic research and language description on the one hand, and child language acquisition on the other hand, because for us—in order to come up with the declarative generalizations about what the native speakers are doing, we do need negative evidence.

So I'm defining positive evidence as suggesting that a given hypothesis is likely correct and negative evidence as evidence suggesting that a given hypothesis is likely incorrect. But in practice, the way people have been using these terms 'positive' and 'negative evidence,' not just in child language research, but in science in general, are much more complex and muddied.

Most commonly, the term 'positive evidence' is used for empirical support of a proposition of the form 'There is an x such that $P(x)$'—so this is a proposition that you can verify by observing a single instance of an x that is P. As soon as you find one example, you know that this proposition is true. That's what people typically talk about when they're talking about 'positive evidence.'

On the other hand, 'negative evidence,' that's what people mean when they are talking about verifying a proposition such as 'There is no x such that $P(x)$'—or 'There is no x that is P.' The verification of this one is a lot harder, because you need to inspect every possible value of this variable x to see whether there isn't one that is P.

These are existential quantified statements for those of you who have a background of logic. What about universally quantified statements such as 'All x are P'? Let's say, all verbs of a given type in this language have a certain feature, something like that. Those are closely related to negative existential statements because for every statement of this kind, there is an equivalent statement 'There is no x that is not P.' So for both negative existentially

quantified statements and for universally quantified statements, it's the case that there is a large—possibly indefinite or infinite—domain of cases that would need to be inspected in order to strictly verify that this proposition is true. In other words, you need a lot more evidence to verify this 'All *x* are *P*' or 'No *x* is *P*' than you need to verify 'There is an *x* that is *P*,' and that's fundamentally epistemologically the problem of negative evidence.

So once again in language acquisition research, it seems to be generally possible for the child to get by on the basis of positive evidence alone. However they do it—it's an amazing feat to be sure. But in language description, generally we're interested in universally quantified statements, or at least implicitly universally quantified statements. So we're interested not just in stating individual observations of events; we are interested in statements, generalizations, about what the speakers of this language do in general, what all speakers do whenever they encounter a particular condition. Or we're interested in generalizations over all tokens of a particular type of a lexical category, a given verb, or noun, or what have you, and so on and so forth. So in other words, one way or another, the bulk of the statements that a language description consists of are implicitly or explicitly universally quantified, or negative existential, which means they require negative evidence.

So basically when we're trying to get from extensional data to the underlying senses, the conceptual representations in the mind of the speaker, we have to filter out pragmatic meaning components from semantic meaning components. So in other words, we have to distinguish between pragmatic and semantic meaning components, and we are going to see that that in particular requires negative evidence. And that's what I'm going to talking about for most of this lecture. And that then will lead us to the semantic properties that pertain to the sense level, and not just to the extension, the set of possible referents.

Now I'm going to take a step back, and I'm going to talk briefly about the relationship between semantic and pragmatic meanings. So this is background.

Let's say you have a classification of different types of meaning. Traditionally, people have been distinguishing between word meaning, sentence meaning, and utterance meaning. More precisely, word meaning is lexical meaning. Sentence meaning is compositional meaning, so that's not necessarily just sentence meaning, but it may be the meaning of phrases. In principle, it's the meaning of any complex expressions that are licensed by the morphosyntax, the grammar of the language. Of course on a constructionist view, the grammar of the language consists of constructions and there is no fundamental difference between constructions and lexical items. So another way of talking about this contrast is in terms of that between lexical items that have a

particular phonological string associated with them and lexical items that are more abstract patterns or templates of combining words into larger expressions, a.k.a. constructions.

Utterance meaning is more precisely speaking pragmatic meaning. It's meaning that occurs at the speech act level, but what exactly that means we have to take a closer look at. Whereas lexical meaning and compositional meaning are studied in the field of semantics, pragmatic meaning is studied, well, in pragmatics.

Now the fundamental difference here is between linguistic expressions and utterances. So words and sentences are linguistic signs, which means they are abstract units of linguistic code, symbolic units of which the speaker has some knowledge, and which you may think of as abstractions over the actual verbal behavior of the speakers. But then they would be abstractions that also occur as abstractions in the speaker's mind, as objects of the speaker's procedural knowledge.

But you can think in principle of these things as entities in some sort of abstract platonic space of ideas. Or you don't, depending on your philosophical inclinations. But whatever they are, they are different from utterances. Utterances are actions. Utterances are things that speakers do. Utterances are things that happen at a particular moment in time, in a particular place in the world, whereas, you know, words and phrases and sentences are not things that have a place or time associated with them. In that sense, they are abstract. Utterances are particulars. Words and sentences are abstract objects.

Let's take a quick look at some examples. *Stunned professors and students saw the fruit bat escaping through the classroom window*. This example is so weird and complex because I made it up so that I could illustrate a whole lot of different problems with respect to it.

So lexical semanticists might want to say that *stunned* is a polysemous verb that includes a concrete sense that has something to do with dulling the senses of a sentient being or something like that. Then it has a metaphorical transferred sense that talks about surprise. A lexical semanticist may be interested in the semantic roles that the verb *see* entails, such as the 'experiencer' and the 'stimulus' role. Lexical semanticists may be interested in the compound *fruit bat* and *classroom window*. *Fruit bat* has a taxonomic structure that names the subordinate species of the species lexicalized by the head, which is *bat*. So that makes it a determinative compound. *Classroom window*, on the other hand, has a part-whole structure, where the head refers to a part of the classroom.

A compositional or formal semanticist, or somebody studying sentence meaning, might want to talk about the referents of the noun phrases *professors* and *students*. Both of these are bare plurals, and since they are bare plurals

made from count nouns, they refer to a number of individuals greater than one. The *fruit bat* and *the classroom* refer to exactly one individual.

The verb phrase *saw the fruit bat* denotes a single event, which happened in the past of utterance time, as indicated by the past tense inflection. The gerund *escaping* refers to a single event, which because of the morphological form here [pointing at screen] was cotemporaneous with the observation, the *seeing* event, and so on and so forth. It goes on and on. There is an entailment here that the fruit bat was visible to the professors and students, and most interestingly, the sentence has several syntactic ambiguities. So, for example, under one reading, both the professors and the students were stunned. Under another reading, it was only the professors who were stunned, but not the students. Also, what the professors and the students observed may have been the escaping of the fruit bat through the classroom window, but it's also possible that what the professors and the students observed was the escaping event, but it may have been the professors and the students' observation that proceeded through the classroom window. All of these different interpretations are associated with different syntactic parses of this string.

Finally, what about pragmatics? This is where we're getting interesting for our purposes, because remember where we headed is trying to sort out semantic from pragmatic meaning components. So a pragmaticist may want to say that the example utterance presupposes that there were students, professors, and a fruit bat, and a classroom window, and so on and so forth. And all the definite descriptions in this sentence carry existential presuppositions.

And of course how do you know these presuppositions? They are presuppositions because if you negate the sentence, or you turn it into a polar question, or you turn it into the antecedent of a conditional, these implications remain intact. So, these are all assumptions that we need to make in order to make sense of this sentence, regardless of the speech act in which it is used. So by virtue of that, these meaning components are presuppositions.

Then come the conversational implicatures. That's really the bulk of what we're concerned with when we're trying to get at the underlying sense, when we're trying to isolate the sense in the extensional data that we've collected. So what are conversational implicatures? These are inferences that aim at getting the most information out of the utterance. So the addressee will infer more than the speaker actually says, and the speaker will be able to rely on some of these inferences. The speaker will be able to predict and rely on the hearer deriving some of these inferences, and therefore the speaker may leave a great deal of the intended information that they're trying to get across to these inferences.

On the other hand, the speaker may also anticipate unwanted inferences, and therefore change the utterance in order to prevent these unwanted inferences, such that if the utterance is finally produced in a given form, the addressee, the hearer, is invited to assume, implicitly, that all the inferences that are still triggered by the utterance, *default inferences*—inferences that are easily predictable for both the speaker and the addressee—that all of these are in fact intended by the speaker.

So, for example, the inference that the professors and the students were inside the classroom rather than watching from the outside, or that the bat was flying rather than walking, or that the bat escaped exiting this classroom rather than flying in, that the window was open rather than the bat crashing through it—all of these are stereotype inferences. They are inferences to the effect that there is one particular interpretation here that is much more prominent than the others based on world knowledge. This in turn depends on the cultural background of the speaker and hearer. And the hearer is generally invited and licensed to assume that all those stereotype inferences go through, because it's incumbent upon the speaker to prevent stereotype implicatures that don't apply. So this is described in the framework of implicature theory that was developed by H. P. Grice, first published in the mid 1970s.

And finally, there's the speech act level. The speaker, in this case, makes an assertion, which means that the speaker commits themselves to the truth of the report they deliver, and that in turn has certain felicity conditions. For example, we are entitled to assume that the speaker is sincere and they actually themselves believe in what they are saying.

So you could say: *It is hot today; but actually I don't believe that it is hot today*. That's not logically contradictory. It is perfectly fine to say that, except pragmatically it's infelicitous. Because when you say *It's hot today*, you are thereby committing yourself to the truth of this utterance. And if you say that you don't believe it in the very next utterance, then that commitment becomes pragmatically infelicitous, so the hearer is at a loss as to what's the point you actually try to make.

So those are three out of four meaning components that are traditionally ascribed to pragmatics rather than semantics: speech acts, conversational implicatures, and presuppositions. The fourth one is indexical or deictic meanings, which are context-dependent such that the lexical form of a word actually taps into the nonlinguistic context and retrieves a referent or instructs the speaker and the hearer to retrieve a referent from the non-linguistic context. I'm talking about expressions such as *this* and *that*, *I* and *you*, *now*, *then*, *today*, *tomorrow*, and so on and so forth. All of these are indexical or deictic expressions that

retrieve components of the speech situation. So that, too, is a type of meaning that's traditionally studied in the field of pragmatics.

You can see to some extent how all of these different meaning components crucially go beyond the word- and sentence-level and depend on the actual utterance. In some sense we could say that all of these are context-dependent in a broad sense, and that fundamentally, since the context is a property of the utterance rather than the abstract sign, the abstract semiotic unit, the sentence or the word, the phrase etc. Words, phrases, etc., since they don't have a place in time and space, they have no contexts. But utterances, even if they consist of a single word, utterances always have a context. Therefore, all of these meaning components occur at the pragmatic level, at the utterance level, and not at the semiotic, at the sign level, at the level of abstract sign.

Expressives would be a fifth component that have recently been attracting a lot of attention again, and so this includes stuff such as *My damn computer is on the fritz again!*, where, you know, this conveys an attitude of the speaker that does not contribute to the conditions that make this utterance true or false. The same goes for honorifics. I have a German example here, *Du bist dran*. It has the solidarity form of the second person pronoun, which is used for people who are your social equal and who you are familiar with. *Sie sind dran* means exactly the same thing, both translate as 'It's your turn' in English, but this one uses the polite form of the pronoun, which is used for people who are socially superior and/or are not solidaries, meaning you are not familiar with them. Of course languages such as Mandarin have much more complex honorific system than this. But even in Europe, the stuff is widespread, and the English type, where you have no honorific distinctions in pronouns is an exception. So some people consider these kinds of expressions a fifth dimension of pragmatic expressions.

Implicatures and presuppositions are pragmatic meaning components in the sense that they are context-dependent, whereas entailments are semantic meaning components in the sense that they occur independently of context whenever an expression is used. In a strict sense, it is only speaking utterances that have entailments. Lexical items and phrases, the constituents of sentences, contribute to these entailments, but don't have entailments themselves. So in the sense that implicatures and presuppositions are context-dependent, whereas entailments are not, what we have to do when we're trying to get from extensions to senses is we have to distinguish between entailments and those pragmatic meaning components, implicatures, and presuppositions.

So here are some definitions. Entailments are relations between propositions. One proposition entails another proposition if there is no possible situation in which the first is true but the second is not. A proposition conversationally

implicates another proposition if and only if the assertion of the first proposition suggests the second in an appropriate context, but doesn't actually entail it. And we then say that the second proposition is a defeasible inference from the assertion of the first. And the computational metaphors Stephen Levinson used in conversational implicatures are default interpretations.

In particular, Grice distinguished between two types of conversational implicatures: generalized and particularized implicatures. It's above all the generalized ones that are default interpretations. The particularized ones are default interpretations in particular contexts, but they always depend on assumptions about the speaker's communicative intention. So the famous textbook example, somebody says *It's cold in here*. It may be that they're trying to get the hearer to close the window or something like that. But that would actually go through given an assumption about the current state of affairs, the nonlinguistic context and the speaker's intentions and maybe their attitudes towards the hearer so on and so forth. So this inference does not depend on the form of this utterance or the type of utterance that it is. It does not in any way depend on the signs, the structure of the sentence, the words that occur in it. It's just a function of the context of the utterance. That's what Grice calls 'particularized conversational implicature.'

In contrast, if you say something like, you know, *Floyd ate some of the cookies*, that implicates Floyd didn't eat all of the cookies. This is also an implicature. It doesn't need to be true that Floyd ate only some of the cookies, that he didn't eat all of the cookies. But this does not just depend on the concrete context of this particular utterance. It's actually something that derives from the word *some* that's used in here. So it depends on the linguistic expressions that are used in the utterance. And in that sense, it is a generalized conversational implicature.

These things in particular, the generalized conversational implicatures, are what Levinson calls 'default interpretations.' Because they are context-dependent in the sense that the context can cancel or block them, but in the absence of such blockage, they will always arise on the basis of a particular trigger, a particular lexical expression, or a particular syntactic construction that invites this inference. In that sense, they are default interpretations. And that's in particular what we need to try to identify in the extensional data that we've collected.

Finally, there is the property of projection, which is the property of presuppositions, although there is a larger class of semantic inferences that have this property. A proposition projects another proposition if and only if any speech act involving the first suggests the second, but the first doesn't entail the second. So the crucial difference between this definition and the preceding one is

the formulation "any speech act involving the proposition," whereas conversational implicatures only arise under assertion.

So this is the property that we've seen before. If I say *Sally chased the squirrel out of her yard*, that presupposes that there was a squirrel in her yard. If I say *Sally didn't chase the squirrel out of her yard*, it still presupposes that there was a squirrel in her yard. If I say *Did Sally finally chase the squirrel from her yard?* It still suggests that there was a squirrel, and so on and so forth. So the independency of the speech act that's built around the particular proposition, that's what sets apart projection, and that's what identifies this property of presupposition.

Here is a classification of linguistic inferences: entailments, implicatures, and presuppositions. Entailments are not defeasible and they are not projected. Implicatures are defeasible, but they are not projected. They only occur with assertions. Presuppositions are also defeasible, although when they are defeated, a truth value gap arises. So if you say *Sally chased the squirrel from her yard*, and it turns out that there never was a squirrel in her yard, there are philosophical reasons. Our first intuitive response would be to say the assertion, the claim, was false. But actually it's a little more complicated than that precisely for the reason that in that situation, the negation *Sally didn't chase the squirrel from her yard* would also be false. But it is not possible for a proposition and its negation both to be false, one of them has to be true. So therefore, strictly speaking, the result here is something that's uninterpretable. We cannot say it's true; we cannot say it's false, it's uninterpretable. But what sets presupposition apart from the others is that they are projected, so they are independent of the speech act.

So how do you test for entailments? And of course this is tricky stuff. Because strictly speaking, the way entailment is defined, right, a proposition entails another if and only if the set of situations or possible worlds or contexts, whatever you want to call it, in which the first is true is a subset of the set in which the second is true. That would mean you have to inspect a very large, possibly indefinite or infinite, set of possible situations in which this kind of proposition could be used. Which is a possible task, and that gets us back to the problem of negative evidence.

In practice, what do you do? One way you can go about this is you ask the speaker: 'Can you imagine any situation in which the first proposition is true but the second is not?' So for example, you want to know whether *The fruit bat escaped from the lab through the window* entails *The fruit bat flew out the window*. Of course we already know that that's not an entailment. We know that it's a stereotype implicature because the prototypical manner of motion for a bat is flying. Therefore, this is, you know, in the absence of further

information, the most likely assumption the hearer will make. But it's not an entailment.

So we ask the speaker: 'Alright, supposing a situation in which somebody says something, can you imagine that to be true even though it's not the case that the bat flew out of the window?' And this is again the kind of task for which you would need a consultant with some imagination, because they have to think about this, and they have to have the ability to extract themselves from the concrete situation and imagine something quite unreal that could be the case. And you know, like I said before, that's not everybody's cup of tea.

Another way to go is to give the speaker a proposition that triggers the putative entailment and the negation of the putative entailment, and see whether the result is a contradiction for the speaker or not. So you tell them, 'OK, tell me whether the following sentence is possible or whether it's a contradiction: *The fruit bat escaped from the lab through the window, but it did not fly*.' And presumably the speaker would tell you 'Yep, that's fine, I can imagine a situation in which this is the case, in which the bat actually climbed up the window and hopped out.' So that's how you prove that it's not an entailment. So you've defeated an implicature. You've identified an implicature by defeating it, meaning by showing that it's defeasible, by showing that it can be negated without making the sentence false.

But the way you did this is by tapping into the speaker's imagination. You got the speaker to help you along, because without the speaker's help, you would be confronted with the impossible task of inspecting an infinite number of situations, something you just literally cannot do. But what the speaker did for you when they helped you out, when they gave you the helping hand, when they tapped into their imagination, is they provided you with a piece of negative evidence.

So here is the corresponding contradiction test. *The fruit bat escaped from the lab through the window. It climbed up to the windowsill and jumped out*—is that possible or is it a contradiction? And the speaker will most likely tell you 'That's perfectly fine, sure, I mean, you know, bats can do that'. And of course the contradiction test is a lot easier than the imagination test. Because, in the contradiction test, you're giving the speaker the scenario, so they don't have to use their imagination. Well, they have to use their imagination to find out whether this is something that's really possible, but they don't have to come up with the scenario themselves, and in that sense, they don't need to use their imagination. So generally speaking, in my experience, tests that rely on contradiction are a lot easier, a lot more feasible to use in the field, than tests that rely on the speaker coming up with a particular situation in which something is or is not the case.

Of course testing for entailments and testing for implicatures in a sense is the same thing. In the imagination test we show that the fruit bat flying is not an entailment. It is not an entailment because it is an implicature. So by showing that this particular inference is defeasible, we've done most of the work for showing that it's an implicature. All we have to do, all that's left to do is just to show that it's not projected, meaning that it is speech-act-dependent.

But if you look at it from the perspective of identifying implicatures, what you want to do directly, to directly test for implicatures, is to show that cancellation does not result in contradiction or that the inference does not arise in a context that blocks it, meaning a discourse context that contains the negation of the propositional issue. So for example, you could try *Despite its injury which prevented it from flying, the fruit bat escaped from the lab through the window.* So here you are constructing a context that explicitly denied the putative implicature that the bat flew. And if that does not result in contradiction, then you know that you've identified an implicature.

So we've already talked about how to identify presuppositions. It basically means you have to show that the inference in question does not depend on a speech act. So if you say *The fruit bat escaped from the lab through the window*, it may give rise to the inference that the bat flew, but if you say *The fruit bat did not escape from the lab through the window*, then there is no such inference because there is no information that the bat did anything at all.

Similarly, the question *Did the fruit bat escape from the lab through the window?* or the conditional *If the fruit bat escaped from the lab through the window*, they don't invite this inference that the fruit bat flew. Therefore, this inference is in fact speech act dependent. Therefore you know that this inference is not projected. So you know that we are not dealing with a presupposition.

I just want to say—I'm surprised that I don't have that information on the slide—there is a forthcoming paper by Judith Tonhauser and colleagues which will appear in *Language*, shortly I hope, which discusses in detail methods for testing for presupposition in fieldwork, under field work conditions. It is a very creative, very resourceful study, which I highly recommend. I think this will be a big boost to people doing research on presupposition under field conditions.

Now that we've hopefully got some sort of understanding of what pragmatic meaning components are, and how they differ from semantic meaning components, let's put that to use in analyzing extensional data and trying to get at the underlying senses, the underlying conceptual representations, in the mind of the speaker.

So to recap: the extension of an expression, it's the set of possible referents. Therefore it's in principle observable, although the observation may not exhaust the entire extension of an expression, but it may give you at least

elements of the extension. The sense, on the other hand, is the underlying concept or representation in the mind of the speaker, and as such it's not directly observable.

This is an example that we already talked about yesterday briefly. It illustrates, first of all, how the primary data in field semantics and any kind of empirical semantics is observational extensional data, in this case, responses by two different speakers to the BowPed task, the topological relations picture series. Of course, these are not just the responses themselves. This is a processed form. These are Venn diagrams. So you see some of the pictures. You see encircled areas around these pictures. These are Venn diagrams. Every encircled enclosed area represents a set of scenes that the speaker uses the same preposition in response to.

Recall the difference between the two speakers. One has the super large use of the preposition *en* for most of the scenes in this domain, whereas the other has the same extension of *en* in the other speakers' usage broken down into a number of much more specific categories which are each labeled using specific, semantically richer prepositions. What accounts for the difference between these responses is not a semantic difference in the mental lexicon entry for this preposition *en* or any of the other prepositions. Rather, it's a difference in the extent to which the speakers make use of implicatures. One speaker is only using *en* for those scenes where it's most informative, thereby relying on scalar implicatures to the effect that any use for which there's a more informative alternative does not apply.

Presumably, the child can figure out how to get from the extensional data she observes to the underlying representations in the mind of the speaker—the competent speaker, the adult speaker—on the basis of positive evidence alone, although we don't quite know how she does that. But the linguist cannot. The linguist needs negative evidence, and that has to do with the fact that the linguist is trying to formulate certain generalizations over the observed data that require negative evidence.

So we know that in principle what we're getting in the extensional data is the product of a number of different components. I'm going to give you a heavily simplified formula, which has all sorts of problems with it, but it will hopefully give you a basic idea. We can write 'Sense + x = extension', where x comprises the effect of implicatures, presuppositions, and semantic transfer.

So implicatures, presuppositions, and semantic transfer all contribute to the observable extension of an expression, and they do not necessarily form part of the sense, although semantic transfer—metaphor and metonymy—may form part of the sense. They just don't do so necessarily. We'll see an example of that in a moment. So it's not true that what distinguishes sense from extension

is solely utterance meaning, solely implicatures, and presuppositions. In addition, semantic transfer can also play a role.

These all have slightly different effects on the extension, on the observed extension. Generally, implicatures and presuppositions shrink the observed extension, meaning they make the extension appear narrower than it actually is based on the sense, the underlying semantic meaning of the expression alone. In other words, implicatures and presuppositions in a sense enrich the interpretation of the expression. That's in particular true of implicatures. On the other hand, semantic transfer has the opposite effect. Semantic transfer widens the observed extension. It makes the observed extension appear larger than it is on the basis of the sense alone.

Let's look at some examples, starting with stereotype implicatures. If you hear *The boxer married the secretary*, this is an example I like to give to my students, and I like to let them figure out all the possible scenarios where this can be true in terms of distribution of gender. And of course, the first thing everybody thinks is 'OK, the boxer is male, the secretary is female.' Then they realize that actually it may also be the case that it's the other way around. The boxer may be female. That's not a stereotypical instance, but it's possible. The secretary may be male, that's also atypical, but it's possible, and of course you can see how this depends entirely on cultural conventions. Depending on whether we're talking about a context in which homosexual marriage is possible, you may also have the scenario under which both are males or both are females, and again you can see the role of cultural conventions.

You see how this makes the extension of this sentence, meaning the set of possible state of affairs that it describes appear, smaller or narrower than it actually is. Because in reality, the sentence can possibly refer to four different scenarios: male—female, female—male, male—male, and female—female. But due to stereotype implicatures, that boils down to just one interpretation that people are most likely to walk away with, which is the interpretation 'male boxer, female secretary.'

In one study on Yucatec, 'path' verbs and the question whether they actually entail translation of motion of the figure in space. I showed you a clip where you have this plank sliding under the little ball and the cylinder here. The question is, was it possible in this scenario to say the ball ascended to the top of the slide? And the Yucatec speakers would say no, unless you actually defeat or block in context the stereotype implicature that it is the figure that moves. So in other words, if you look at the description, the Yucatec counterpart of the description 'The ball rolled up the slide'—if you look at the Yucatec expression of this proposition, the stereotype implicature suggests that the figure moves, and that narrows the observable extension of this description to situations in

which the figure moves, thereby excluding this scenario. But once you add a little context that blocks the stereotype implicature, all of a sudden the description becomes acceptable in reference to this scenario. And it turns out that it does in fact fall inside the extension of this description, which in turn means that the description does not entail figure motion.

Another very important kind of conversational implicatures that have an impact on the data we observe in the field are manner implicatures. Manner and stereotype implicatures are complementary in the way they work on a sort of division of labor basis, if you know what I mean. So stereotype implicatures pick up stereotypical implications, inferences from expressions that are simple. Manner implicatures do the opposite. Manner implicatures kick in when you're using an expression that's more complex than the simplest possible alternatives or that uses lexical expressions that are rare, infrequent, or from an unusual register, and so on and so forth. So there is something not obvious about the expression that you're using. There's something out of the ordinary, and that suggests that the most stereotypical scenario does not apply.

A very well known example of that is the division of labor between different causative constructions in terms of constructions that invite inferences to direct causation and constructions that invite inferences to indirect causation. This was first described by James McCawley in the 1970s. So imagine you have *Sally stopped the car* versus *Sally caused the car to stop*, and you ask people, 'Assume that in one case Sally hit the brakes, but in which one? And in the other case Sally did not, but in which one?' In the other case, Sally did something more unusual. Let's say Sally actually was not in the car. She was not driving herself, but she may have stepped into the street in front of the car causing the driver to brake.

People will tell you that that interpretation is the direct causation scenario of the simple causative verb, whereas the light verb construction, the periphrastic, the syntactical causative gets associated with the unusual interpretation such as Sally stepping in front of the car, the indirect causation scenario. It's unusual in the sense that it's not the simplest causal chain that you could come up with to interpret this sentence.

My colleagues and I did a study of the semantics of causative constructions in a bunch of different languages using video clips. What we found was that, for example, whenever there was a causal chain that showed a spatial gap or a certain lapse of time between cause and effect—so imagine you see somebody hit a plate with a hammer, but instead of the plate going to pieces immediately, that only happens after a moment passes by—whenever that kind of thing happened, the Yucatec speakers would go to a periphrastic causative, something like *cause to stop*, rather than to use a simple causative verb.

We get a very similar effect in the Cut and Break study. We found that speakers from a number of languages that have various kinds of complex predicates, such as Lao and Sranan, both of which have serial verb constructions, and Yucatec Maya, which has verb-verb compounds, which are sort of the polysynthetic language's equivalent of serial verb constructions, or German, my native language, which has verb-particle constructions. In all of these languages, as soon as you show speakers a kind of scenario that's unusual in terms of the combination of the object that's being acted on and the instrument that's used to act on them, you get a complex predicate.

So for example, take this clip, which involves a guy severing a piece of cloth with a mallet. Then you get complex responses, complex predicates. One of the German speakers said something like *Er zerhämmerte Omas Kleid* 'He *zer*-hammered grandma's dress,' and *zerhämmern* is not a verb that you're going to find in any dictionary of German. He just made that up, the speaker made that up on the fly. But it illustrates the need to use a complex predicate when you're encountering a non-stereotypical scenario.

Similarly, a Yucatec speaker, in response to the same clip, produced a verb-verb compound. He said 'He rip-hit the piece of cloth with the hammer,' meaning 'He hit it, causing it to rip,' or rather, actually what the construction really means after a lot semantic analysis is 'He hit it in a manner so as to rip it,' but that's a secondary issue.

So the point is you get a distribution, a sort of division of labor where the more complex predicates are used to pick up the less typical, the unusual combinations of instruments and themes, and get simple causative verbs to take care of businesses as usual, dealing with the stereotypical instrument-theme combinations. Again these are extensions that are narrower than the semantics of the particular verbs, and verb compounds, actually allow for, due to the influence of manner implicatures.

Scalar implicatures, that's the preemption effect we've already talked about. That *Floyd ate some of the cookies* suggests that *He didn't eat all of the cookies* but does not entail that. *Sally started writing a book on semantics* implicates *She didn't finish the book*. I want to give you a quick example of how that mattered in my fieldwork. It turned up in an analysis of the demonstrative system of Yucatec. So Yucatec, strictly speaking, doesn't have demonstratives of the kind that we know from languages such as English. Instead it has clause-final particles that are triggered by the determiner, and there is a particle *=a'* which is used for things proximal to the speaker, and there is a particle *=o'* which is used for things that are distal to the speaker.

Now the question is this: is it actually the case that semantically, *=a'* means proximity and *=o'* means distance from the speaker (or from the speech

situation)? First of all, this was never particularly likely, because the particle *=o'* cannot only be used for distal themes, but it can also be used for reference to objects that were mentioned before in discourse; meaning it's not just used deictically, exophorically, with respect to the speech situation, picking up a referent in a speech situation, it can also be used anaphorically, which the particle *=a'* cannot.

So my hypothesis was that as a matter of fact *=a'* is positively semantically specified for exophoric reference to objects that are in the proximity of the speaker in the speech situation, whereas *=o'* has a much more general meaning. It's a general indexical, it can be used exophorically or anaphorically and in fact in other ways as well. But it gets associated with the distal domain simply by preemption, because *=a'* is the more semantically specific, and therefore more informative, partner in this opposition.

And of course that suggests that is in fact possible to use *=o'* for the items that are in the speaker's own proximity as well. I was able to show that that's the case. Although, strictly speaking, you cannot use *=o'* for things that are on the body of the speaker, which is a very interesting phenomenon, and it's not easy to explain.

Let's look at a quick example for how presuppositions mattered in my fieldwork. I mentioned earlier that in my dissertation, I tried to show that Yucatec is a tenseless language, meaning there is no tense marker in this language. Yucatec has a very rich aspect-mood marking system, which I described in detail, and that system also includes what looks like markers of metrical tense.

Metrical tense markers are tenses that are restricted, for example, to events that happened earlier on the day of utterance, or to events that happened on the day before the day of utterance, or to events that happened long ago, or events that the speaker predicts will happen a long time from now, from the moment of utterance. Obviously, if that's what these things are, metrical tenses are tenses. Therefore, they will contradict my claim that Yucatec is a tenseless language. But I was able to show that while these expressions, such as this recent past marker *sáam*, which generally speaking is used for events that happened earlier on the day of utterance. These people semantically only talk about distance from a reference point, but not actually about the temporal relation between the time of the event or the reference point. That relation is actually presupposed rather than coded. It's not part of the semantics of this expression.

And how do you show that it's presupposed? Well, suppose you negate a sentence in which the marker in question occurs. Then of course, due to the speech act independence of presuppositions, we expect that the presupposition remains intact, and that is exactly what we observe. So if I say 'The

combi'—I have to explain this—the *combi* is what mainstream Mexicans, for want of a better expression, would call a *colectivo*, it's a collective taxi. In this Mayan village which I have been working in for a long time, people call that a *combi*. So if I say *Sáam sùunak le=kòombi=o'*, that means that the *combi* returned a while ago, earlier on the day of utterance. Now if I say (9), I can continue this by saying (a), but it would be infelicitous to continue with (b).

(9) *Ma' sáam sùunak le=kòombi=o',*
NEG REC turn\ATP:SUBJ(B3SG) DET=van=D2
'It's not a while ago that the bus returned; …'

a. *… inw=a'l-ik=e' h-ts'o'k mèedya òora.*
A1SG=say-INC(B3SG)=TOP PRV-end(B3SG) half hour
'… I think it was half an hour ago.'

b.?? *tuméen ma' sùunak=i'.*
CAUSE NEG turn\ATP:SUBJ(B3SG)=D4
'… because it hasn't returned yet.'

So the fact that the *combi*'s return precedes the reference time, which in this case is the moment of utterance, is not actually a semantic property of this construction; it's a presupposition. Which, again, has the effect of making the extension of these morphemes appear narrower than they are, based on the semantic properties of the markers alone.

Semantic transfer, metaphor and metonymy, of course may be a part of the mental lexicon entry of the expression itself. That's the case if it's conventionalized, but it may also be something that speakers create on the fly, such as in the case of the famous Nunberg metonyms, as in *The ham sandwich is at table 7*,[1] where the speaker is supposed to be a waitress talking to another waitress, and what she means by *ham sandwich* is not actually the sandwich, but rather the person who ordered it. That's a metonymic reference from the sandwich to the person who ordered it. It's a kind of semantic transfer, but clearly this is not a conventional meaning of the noun, the compound *ham sandwich* in English. It's something that the speaker here does on the fly for this particular utterance.

This can happen with metaphor as well. That's something I've been studying as part of a study of meronymy, meaning expressions for object parts, something that is very prominent in Mesoamerican languages and has an important

1 Nunberg, Geoffrey. 1995. Transfers of meaning. *Journal of Semantics* 12: 109–132.

influence apparently on strategies of spatial reference in Mesoamerican languages.

Now in order to study these kinds of expressions some more, we had a set of novel objects made for us, which are objects of a shape that's unfamiliar to both speakers of Mesoamerican languages and to other people from outside that area as well. So they don't have any conventional interpretation. And the question was, would people be able to refer to the parts of these objects without first establishing some sort of overall interpretation and also an orientation of the object? There are some proposals in the literature for some of these systems that that should be possible. So that's what we wanted to test.

What I found for Yucatec in particular was that it depends a little bit on what parts we're talking about. So there is a big difference between surface parts and volume parts, or strictly speaking, surface parts and parts for edges or extreme points on the one hand, and volume parts on the other, because when it came to volume parts, generally speaking, the identification did rely on people establishing an interpretation of the overall object, and it also was the case that, very frequently, similes and hedges would accompany the use of terms for volume parts.

So they would, for example, say 'The little ball, it is as if it had *bèey kan-p'éel yòoka*,' as if it had four legs. So we're not actually saying that the parts of this object can be readily referred to as legs in a way that's going to be obvious to the interlocutor, to other speakers of the language. Rather we're going to have to explicitly mark the fact that we are treating these as legs. So this is very clear, overtly metaphor. It's overtly marking semantic transfer.

Or, for example, '*U mehen ba'lilo'b dée mehen òoko'bo*,' 'it's little things that are sort of like legs,' and here the speaker said explicitly 'let's call that its arms, let's call those things its arms.' So this was in a referential communication task where the people again had a screen between them and their goal was to instruct each other to put little bits of play-dough onto these different parts; in order to do that they have to successfully identify these parts verbally.

So this is what they had to do when it came to volume parts, whereas when it came to surface parts, so terms such as 'back' and 'front', 'top' and 'bottom,' so on and so forth, they would never ever use any similes or hedges. So it turns out that the surface and edge parts have more general, more abstract geometrical meanings, which allow the speakers to apply them to any arbitrary object regardless of interpretation, whereas the volume parts are labeled on the basis of body-part metaphors, and these body-part metaphors are to some extent conventionalized.

So you noticed that the effect of this kind of semantic transfer really is the inverse of the effect of the pragmatic meaning components. The pragmatic

meaning components make the extension appear narrower than it is on the basis of the semantics of the expression alone, whereas semantic transfer does the opposite: by adding elements, such as the parts of these novel objects, to the extension that actually don't fall under the extension purely based on the sense taken as a static concept alone.

The empirical basis of field semantics is inferring senses from observed extensions, since semanticists aren't mind readers. To do this, semanticists require negative evidence, meaning evidence that goes beyond the observed extension and talks to what can be in it or what cannot be in it. And to achieve this, semanticists manipulate real or imagined situations and observe how this affects native speakers' intuitions about the applicability of certain expressions in reference to these situations. The process of isolating senses from extensions involves the distinction between semantic and pragmatic meaning components and that between literal and transferred meanings.

Observed extensions are narrowed by pragmatic enrichment but widened by semantic transfer. And the core phenomena of semantics and pragmatics, such as entailment, contradiction, ambiguity, anomaly, implicature, presupposition, and speech act meaning, can be explored in the field directly or indirectly on the basis of native speaker intuitions for conditions of successful reference or truth conditions, which is not to say that these phenomena can be captured exhaustively in referential terms. And diagnostics of lexical-semantic relations are always based on intuitions for either contradictions or anomaly. Well, I should say Alan Cruse mentions a possible third type of evidence, which has much more reduced currency, but I am a little bit dubious about that. It is intuitions for analogies, analogies of semantic relations across pairs or triplets, etc. of lexical items.

LECTURE 5

Ethnosemantics and Cognitive Anthropology: a Short History

It's actually very appropriate to be in a forestry university for the topic of this lecture, which is partly going to touch on ethno-botany, meaning the cultural and culture-specific linguistic categorization of plants and animals. Besides, I actually do field work literally in the forest, in the jungle, in the bush.

So we are making our transition from the first topic of these lectures, field semantics, to the second topic, which is semantic typology. We are going to talk about ethnosemantics, which is, if you will, a version of semantic typology, but one that as a research program was conceived mainly not by linguists, and not with an explicitly typological focus. It was conceived by cultural anthropologists and linguistic anthropologists and specifically in response to the Linguistic Relativity Hypothesis.

One important intellectual background of the beginnings of the ethnosemantic research program is the view of linguistic semantics that was prevalent in Structuralism, the approach to the structure of language in the first half of the 20th century—basically the beginning of modern linguistics as we know it. The structuralists had a conviction according to which the semantic categories expressed in language are purely language-specific and culture-specific. The only factors that would influence what kind of properties, what kind of stimuli, get classified together in a given category would be the structure of linguistic code itself and questions of cultural utility.

In other words, what was absent from the reckoning of most structuralists was the view of the human mind as a universal principle that would structure linguistic semantics culture-independently, something that most people nowadays take for granted, but that presupposes a view of the mind, a view of cognition that did not arise, especially in North America, until after World War II.

All original audio-recordings and other supplementary material, such as any hand-outs and powerpoint presentations for the lecture series, have been made available online and are referenced via unique DOI numbers on the website www.figshare.com. They may be accessed via this QR code and the following dynamic link: https://doi.org/10.6084/m9.figshare.11419122

© JÜRGEN BOHNEMEYER, 2021 | DOI:10.1163/9789004362628_006

This view of the semantic categories expressed across languages to be entirely language-specific is epitomized by a famous slogan that goes back to Franz Boas, the founder of linguistic anthropology and one of the founders of both cultural anthropology and linguistics. Boas said "Thus it happens that each language, from the point view of another language, may be arbitrary in its classifications."[1] What he meant by that is, if you look at the semantic categories expressed in one language, you cannot from there predict the semantic categories expressed in any other language. That's what he means by 'arbitrary.'

I have a made-up example that I like to use to make this idea a little more concrete, which is that of an imaginary color term system. Let's consider the so-called Munsell color chart, which I will talk more about today. It's a classification of colors in terms of hues and brightness. It distinguishes 40 different hues and 8 different levels of brightness. So that gives you 320 cells. The divisions that I've drawn on this grid are supposed to represent two areas, except one of them is non-contiguous. This is a complication that has to do with the fact that there are two non-contiguous wavelengths, or bands, that are translated as red, roughly, by the human neuro-physiological system. So there are two different frequency bands of light that are interpreted as red. That's why you have red on both ends of this continuum.

So in other words, what I'm really trying to get at here is the idea of a color term system that distinguishes between two categories: red and everything else, a red/non-red color term system. My point is, from the perspective of a structuralist, you would expect this is a very, very common color term system in the languages of the world. Why is that? Because this system has a structure that the structuralists thought was very efficient and very economical for dividing up a domain. It's a privative opposition with one marked and one unmarked member, and the marked member happens to be culturally salient in many societies. It's the color red, the color of blood, the color of danger, and so on and so forth.

Now let me ask you this: what do you think, how many languages actually have a color term system like that: a red/non-red color term system? As a matter of fact, *no* language has a system like that. Of the hundreds of languages that have been tested to date, which is actually a pathetic sample if you consider the fact that there are between six and eight thousand languages still extant on the planet, there is not a single instance of a red/non-red color term system.

We know that thanks to the efforts of the ethnosemanticists and semantic typologists. Obviously it's a finding that requires an explanation. The

1 Boas, F. 1911. *Handbook of American Indian languages, Volume 1.* 26. Washington, DC: Smithsonian Institution.

explanation has to refer to cognition, which is something that was not on the map when people like Franz Boas started out thinking about linguistic categories. But we wouldn't have gotten to this insight if it wasn't for Boas and the early ethnosemanticists. So today what I'm going to try to do is to retrace the steps of how we got to know what we think we know now about the semantic categories of the languages of the world.

Of course this view that semantic categorization is arbitrary in the sense that it's entirely language- and culture-specific was a very important intellectual prerequisite for the Linguistic Relativity Hypothesis. That's expressed in this wonderful quote from Whorf, who said:

> Just as it is possible to have any number of geometries other than the Euclidean which give an equally perfect account of space configurations, so it is possible to have descriptions of the universe, all equally valid, that do not contain our familiar contrasts of time and space. The relativity viewpoint of modern physics is one such view, conceived in mathematical terms, and the Hopi Weltanschauung is another and quite different one, nonmathematical and linguistic.
>
> WHORF 1950: 67[2]

What Whorf is saying here is (a) languages are arbitrary in their classifications, and (b) therefore we can have entire models of the universe that are language-specific and that have little to do with how the universe is conceptualized in other cultures by the speakers of other languages. Now I've always found this a very, very inspiring idea. It's something that has fascinated me for years and years. But I'm also interested in finding the empirical evidence for and against this view.

So what we are going to do is we're going to take a brief tour through the main areas of ethnosemantic research from the inception in the 1950s mostly through the most recent work that's being done currently. So most of that research has focused on three areas. The first of these is kinship terminology, which is an old topic in cultural anthropology. It's one of the main research strands of cultural and linguistic anthropology since the 19th century. The other two areas are color terminologies and ethnobiological classification. Then we'll discuss some implications of that and in the end look at the new domains that people started working on within the last ten years, although a lot of that research is more explicitly typologically oriented and it's now done primarily no

2 Whorf, B. L. 1950. An American Indian model of the universe. *International Journal of American Linguistics* 16(2): 67–72.

longer by anthropologists but rather by linguists and psychologists. So that's part of the shift, the transition, from ethnosemantics to semantic typology.

So let's start with kinship terminology. A general property of the ethnosemanticists has been that they look at the semantic categorization of the natural world. So they look at domains of conceptualization where you have a good chance of finding a universal basis of categorization, but at the same time you also have a strong motivation for culture-specificity, so that you can pit the two against one another and see which one is the stronger factor, or put differently, what is the effect of each of these two forces.

So in the case of kinship terminology, there is a biological basis to the phenomenon of kinship that we can assume exists culture-independently, although we know that every culture has a different spin, or interpretation, on this biological basis, and in many cultures biology is not a major factor in how people conceptualize kinship. Nevertheless, we could say there's a possibility there of a universal basis. On the other hand of course we know that kinship relations are cultural institutions. They vary a great from culture to culture. Cultural anthropologists have worked for a long time on how institutions of kinship reflect factors such as marriage taboos and descent rules, inheritance rules, and so on and so forth.

Lewis Henry Morgan was pretty much the founder of the typology of kinship term systems that is more or less still considered valid by anthropologists today. It was conceived in the 1870s on the basis of a questionnaire study with data from a bunch of languages around the world. So this really was an early typological study.

Lewis Henry Morgan was a lawyer, and as a matter of fact, many of the early anthropologists who got fascinated by the problem of kinship were professionally lawyers, or lawyers by training. For obvious reasons: lawyers have to deal with matters of inheritance, which hinge on kinship issues. Lewis Henry Morgan lived in my adopted hometown of Rochester, New York. He worked for the Seneca Nation, an Iroquoian people, and also with other Native American peoples, and that's how he discovered that different languages and cultures have very different concepts of kinship. So he wanted to know how much variation is out there and what are the constraints on this variation.

Now let's talk a little bit about how anthropologists and linguists can go about studying the kinship terminology of a given language. Here's an important concept that we'll talk about a lot today: the concept of an etic grid. And I'll try to explain the term in the afternoon; but basically what an etic grid is, is a presumed language- and culture-independent classification of a particular domain that you're going to use as a frame of reference in which you can map the conceptualizations of each culture and language.

So in studies on kinship terminology, the etic grid is traditionally, going back to Lewis Henry Morgan, a network of genealogical relations. You start out from ego—kinship terms are, in a sense, deictic, meaning they talk about relationships with respect to the speaker or the cognizer. Then we have the ego's siblings. The circles are females. The triangles are males. Then we have ego's parents. The equation mark indicates a marital relation. The horizontal lines give you sibling relationships, and the vertical lines give you descent relationships, or offspring relationships.

It is not presupposed here that every culture and language that can be mapped on this grid has to have a biological interpretation of these genealogical relations. That is left blank. In a sense, we are treating these genealogical relations as variables, the interpretations of which can be entirely language- or culture-specific. The idea is merely that every kinship term system recognizes that any person has one mother and at most one father. Whether that's the biological mother and the biological father, or whether the identity of that mother and that father is determined some entirely different way, is left outside the focus of consideration here.

We can use this genealogical grid to map the classifications that each language makes against this grid, and we're going to see that in a moment when I introduce those types of those kinship term systems that Lewis Henry Morgan found and that we still distinguish today. You see how we use colors against the same genealogical grid to illustrate the language- and culture-specific groupings. So that's the idea of an etic grid.

Now how do you get the terminology? The best way to go about this is to work with a large number of speakers, and from each speaker, I mean, elicit that person's, collect that person's actual genealogy—meaning the actual genealogical network, including the sibling and cousin generation, one generation down, and the parent generation—the actual genealogical network of this particular person. Some people in some cultures have two genealogical networks because they are affiliated separately on the mother's and on the father's side, so this can be quite complex.

Once you've collected the speakers' genealogy, you ask them, 'Okay, so imagine you and I, we meet this person here, and you want to introduce me, so you're going to say "This is my X." I want you to tell me what you're going to use for X here.' So it's important to get the label that the speaker would use for reference rather than for addressing the particular kin. Universally, across languages, kinship terms are used as terms of address as well, but the categories are often different. For example, when I was a kid, my parents taught me to address friends of the family as 'aunts' and 'uncles.' You know, they weren't my aunts and uncles in the kinship sense. We address people as 'brothers' and

'sisters' and whatnot. We are all familiar with that idea. So it's important to make sure you get the terms for reference and not those used for address.

Of course there can be a lot of privacy issues that come up when you're trying to get a consultant's actual genealogy. So many people will prefer to work with made-up imaginary genealogies. Which is in principle fine. The only problem with that is, because kinship is based on a network of relationships, it's very easy to get lost in the classifications of a kinship tree. So if you're working with imaginary genealogies, it's much easier for the speaker to get confused about the category of a given kin, whereas if they are talking about the actual kin, they're going to know how they refer to these people.

Let's talk about Lewis Henry Morgan's types. He found that the kinship terminologies in the languages of the world—and he got data from over a hundred languages based on responses from traders and missionaries and so on, all these different folks he sent his questionnaires to—he found that all of these pertain to six different types, meaning, you know, each falls under one of these six types.

The labels he used for the types aren't entirely the same that we use today, but the types are the same. People have over the years suggested additional types, but Morgan's six types are still recognized by anthropologists these days. In fact, there is much more variation than this classification, in terms of six types, suggests. As a matter of fact it's very hard to find two kinship terminologies that are entirely identical. So in some ways you can think of every actual terminology that you find as a mix of multiple of these types.

We'll start with two extreme opposites, the Sudanese and Hawaiian types. The Sudanese is the most descriptive, in the sense that there is a distinct term for each genealogical relation, and the Hawaiian type is the least descriptive system, meaning it's the system that has the highest number of what's called 'classificatory' kinship terms. And classificatory kinship terms are kinship terms that merge multiple genealogical relations. So, English examples are the terms of *aunt* and *uncle*. The term *aunt* doesn't distinguish between the maternal and paternal aunt and similarly for the term *uncle* in English, and so in that way those are classificatory terms.

In Hawaiian type systems, all kinship terms are classificatory. The only distinctions that are made are in terms of generation and sex. So there is no distinction between siblings and cousins. There is no distinction between uncles and parents. So the father term is extended to the father's brothers. The mother term is extended to the mother's sisters, and so on.

I've seen plenty of different Mandarin systems over the years, because in my courses I often assign students the task of collecting a kinship term system

from a language they don't speak and analyzing it, so a lot of people wind up collecting Mandarin systems. They all look different. Up to now I've not seen two Mandarin systems that look entirely the same, but they are all Sudanese. They all fall under the maximally descriptive type.

So next step we have the so-called Eskimo type, which is characterized by the so-called bilateral emphasis, meaning there's a merger of kin across the divide between the mother's lineage and father's lineage. In other words, these are symmetrical systems. So in other words, there is no distinction between the maternal aunt and paternal aunt, no distinction between the maternal uncle and the paternal uncle, no distinction between cousins on the mother's side and cousins on the father's side. This is the system that most modern European languages, including English, use—although English, or Old Saxon, did apparently have a Sudanese system. Similarly, Latin had a Sudanese system, and then the modern Romance languages wind up with Eskimo-style systems, and so on.

Next step we have the Iroquois type, which is characterized by the phenomenon of bifurcate mergers, which means there are extensive mergers but they respect the bifurcation. So these are systems that are precisely *not* symmetrical. So you do merge the mother with the mother's sister, but the father's sister gets a different term. Ditto for the father's brother as opposed to the mother's brother.

The merger extends to the role of linking kin, which means there is a distinction between cross cousins and parallel cousins. So the parallel cousins are the offspring of the mother's sister and the father's brother and the cross cousins are the offspring of the mother's brother and the father's sister. This is important because many cultures have incest taboos against parallel cousin marriage, but at the same time encourage cross cousin marriage.

Finally we have the Crow and Omaha systems, which are, if you will, variations of the Iroquois system, but with a twist. These systems do something funny with the cross cousin terms. In the Crow type, the paternal cross cousins are raised one generation, meaning a person that, from an English perspective, would be a distant cousin is actually treated like a father or an uncle. And the maternal cross cousins are lowered one generation, meaning they are treated on a par with ego's siblings and also the parallel cousins. And the Omaha type is the inverse—a mirror image—of the Crow type, so here we have the cross-generational raising on the maternal side and the cross-generational lowering on the paternal side.

The classic interpretation is that Crow type systems occur predominantly in societies with matrilineal descent, meaning societies in which a child is

considered part of the mother's family but not the father's family. And conversely, Omaha type systems are by hypothesis associated with patrilineal descent, meaning the child is considered part of the father's family only.

Those are the six types that Morgan postulated. Over the years people have suggested additional types. In particular many people recognize a Dravidian type, which is also a variant on the Iroquoian type. But be that as it may, anthropologists have long assumed that there are a small number of kinship term systems, which actually isn't obvious from my perspective. I'm looking at this from the perspective of a semantic typologist. Having seen many actual kinship terminologies, it's not so obvious to me that these traditional six types really play a particularly salient role in kinship typology, if you will. But that's been the traditional assumption.

Now the question is if there are these types, how can we account for them, and how can we account for the type that we find in a given language and among the speakers of a given language, who are members of a particular society? The interpretation Morgan himself came up with is that of social evolution. Consider the Sudanese-type system is the most complex system—Lewis' idea was that that's what you're going to find in the most complex type of society. The Hawaiian system is the simplest system; so this is, according to Lewis, what you're going find in the most primitive society. There is an evolution in stages that leads from the Hawaiian type to the Sudanese type, which I suppose would mean for English that it has gone downhill somehow in social structure with respect to the Old Saxon days.

Obviously—well, I'm assuming this would be obvious for many people nowadays—this is completely bogus. The Hawaiian system, which happens to be the system that you find in many Polynesian societies, is from a highly stratified feudal society, with very complex social organization, similarly for other Polynesian societies. So there is no relation whatsoever between the complexity of the kinship term system and the complexity of the social organization.

This illustrates a very important point about early 19th century attempts at classifying systems of linguistic semantic categorization. These approaches were often marred by assumptions that, from our perspective nowadays, are racist and Social-Darwinist, and that explains probably why after a lot of intensive work on this problem in the 19th century, there was no follow-up for about fifty years. Research on this problem went completely dark, vanished from the map, because you know this stuff had been marred by a lot of very problematic assumptions.

The British anthropologist Radcliffe-Brown tried to come up with an interpretation of these systems along the lines of cultural utility. So he proposed a hypothesis that talks about the relationship between Crow-type systems and

matrilineal descent and Omaha-type systems and patrilineal descent. That is something that Radcliffe-Brown would explain directly in terms of, well, "this is a classification that is particularly useful for societies with matrilineal descent;" or "this is a classification that is particularly useful for societies with patrilineal descent."

And then in the 1950s, Floyd Lounsbury pointed out what he thought was a major problem with Radcliffe-Brown's approach, namely, that there are important mismatches where you have in fact Crow-type systems in patrilineal societies and Omaha-type systems in matrilineal societies. More generally, Lounsbury argued that there is a step missing here, which is, before we try to understand the usefulness of these classifications for the cultures that use them, we should first try to identify the cognitive basis of the categories involved. And in doing so, or trying to do just that, Lounsbury became one of the pioneers of Cognitive Linguistics.

So Lounsbury identified a number of rules that generate the conceptual categories involved in these different systems. And I have an appendix. We don't have the time to go over the rules that Lounsbury proposed, but I do discuss them in the appendix of the handout briefly. But essentially, what Lounsbury's proposal entails is that every classificatory kinship category has a focal instance, a prototype as we would nowadays say, which is a nuclear family kin, the mother or father most typically. Then there is a productive rule that allows the successive addition of non-focal members to this category. So it's a recursive rule. These are recursive rules.

That actually, as far as I know, was the earliest mention in the history of cognitive science of the idea of focal instances. Berlin and Kay took the idea of focal colors and the term 'focal' very much from Lounsbury's work on focal instances in kinship categories and Eleanor Rosch of course developed prototype theory as inspired by Berlin and Kay's work.

Eve Danziger, one of my former advisors, has done some work to test an important implication of Lounsbury's proposal, which is the property of kinship term systems having a structured extension such that the nuclear kin members are better instances of the category than the more distantly related members of the extension. She tested this against data from language acquisition in Mopan Maya, a sister language of Yucatec, the language I've been working on, and failed to confirm it. So what Danziger found is that when Mopan kids learn the classificatory terms of that language, they immediately start out considering more distant kin as equally good instances of the category as they do the nuclear kin.

Of course there has been a lot of debate about the reliance on genealogical network relations as an etic grid for comparing kinship terminologies across

language and cultures. But as far as I'm aware, people have not been able to come up with an alternative. I should say, before we leave this topic, two weeks ago a couple of articles were published in the journal *Science* on kinship terminologies. One by Charles Kemp and Terry Regier, who look at around 500 kinship term systems from around the world based on data that were published by Murdock in the 1940s. They're analyzing these kinship term systems in terms of what you might want to call 'cognitive efficiency,' in the sense that the classification has to be simple enough to make it relatively easy to identify the category any particular referent, any particular kin, falls into, and on the other hand 'communicative efficiency,' meaning, a system is communicatively efficient if it makes it easy for a hearer to identify the intended referent of a kinship term.

And of course you see that these are opposing forces, that there is a tradeoff here in the sense that the most communicatively efficient system would be one that uses a distinct term for each possible kin, but that's also going to be a system that is cognitively least efficient, in the sense that it requires the most complex categories and therefore the categories that are hardest to conceptualize.

What Kemp and Regier found is that the actually existent systems, at least those Murdock described in [1949], tend to all cluster—if you map out communicative and cognitive efficiency, or conversely communicative cost and cognitive cost, in a coordinate system, then of course systems are optimal if they are in that corner where both cognitive cost and communicative cost are minimal, and that's exactly where most of the actually existing kinship terminologies cluster.

I'm not sure how interesting a finding this really is. Honestly, I'm still literally wondering about that. In particular, I'm not sure that it tells much about how the domain of kinship is actually conceptualized across the world. And Kemp and Regier also explicitly suggest that similar principles will govern the lexicalization of any linguistic domain. Nevertheless, probably, this is an idea that we're going to hear more about in the near future.

The other article by the way is by Stephen Levinson, and he points out some possible doubts and criticisms with Kemp and Regier's approach. I also want to mention work that has been done in recent years by Fiona Jordan at the Max Planck Institute for Psycholinguistics in the Netherlands using phylogenetic methods to study the evolution of kinship terminologies, for example in the Austronesian language family, or across languages of Melanesia, and so on. So this is very interesting work, but it hasn't so far produced any results that really radically challenge the way we've been thinking about kinship classification.

Let's move on to color terminologies. Once again we're looking at this from the perspective of pitting nature against nurture, or culture. In the case of the classification of colors, the presumed universal basis, which may or may not affect the actual linguistic categorization, would be the color sense, the neurophysiological processing of color, which as far as we know does not vary across human populations.

This is in fact something that was investigated in another pioneering study in semantic typology in the 19th century by the German ophthalmologist Hugo Magnus. So, at this time, it was already known that some languages had fewer words for colors than others. In particular, the British philosopher and politician William Gladstone—who was at one point British prime minister—pointed out that the classic European languages, like Latin and Greek, had many fewer color words, possibly no genuine color words in the sense that modern European languages do. That engendered a debate about whether the speakers of the language, meaning ancient Roman and Greek people, had a diminished color sense, whether their color conception was different from that of modern people.

Magnus wanted to study this problem, and did so on the basis of a very similar approach to that Berlin and Kay took a hundred years later. So he compiled a set of color chips, much smaller than Berlin and Kay's. He had just ten chips, and he sent it to people who would have access to speakers of some sixty languages, and on the basis of the returns concluded that you cannot, from the complexity of the color term system, conclude that the speakers of these languages have different color perception, which is what is basically Berlin and Kay conclusion a hundred years later. So that was Magnus.

In the 1950s, Roger Brown and Eric Lenneberg did a study on memory for color and how it's potentially influenced by, or aligns with, lexicalization, based on English speakers alone. They believed they had found evidence for a Whorfian effect, although nowadays we would say that in order to show that, you would have to look at speakers of multiple different languages, not just at speakers of English.

Conklin (1955) is an often-cited example for an early ethnosemantic study. That was a study focusing on just one language. And rather than a study with the goal of finding uniformity, it was one that specifically tried to uncover the extent of language-specificity in the color categories of this Austronesian language, Hanunó'o, of the Philippines.

And then came Berlin and Kay's seminal study, and we've already talked about what they did. So they started from the Munsell color chart—the Munsell company is a company that makes hues and dyes—and uses this chart

as a reference system. Berlin and Kay used those 320 plus ten black and white chips—330 chips, and presented them to speakers of 20 different languages. Actually their students did. They did a seminar on this topic at UC Berkeley in the 1960s and had the students work on this problem apparently.

They came up with two generalizations. One according to which the focal instances of these different color terms cluster across languages. Here you see a representation of the grid which shows the 11 possible clusters. So each dot is the focal color of some basic color term in some language. So you have clusters which you can label according to their English names, like 'pink', 'purple', 'blue', 'green', and so on. And according to Berlin and Kay, there are only these 11 clusters. So every possible focal color, meaning every best possible instance of a basic color term in any language has to fall in one of these 11 possible clusters.

Here they are presupposing that every language *has* basic color terms. And in the methods they used, they identified the basic color terms independently of the semantics of these terms, just on the basis of criteria such as monomorphemicity, meaning they are not complex. They are autochthonous terms, native terms rather than borrowed terms. At least synchronically, speakers don't recognize them as borrowed terms. They have to be of general currency, so they are not restricted to particular registers. And they are semantically general in the sense that they will label any manifestation of the relevant hue,

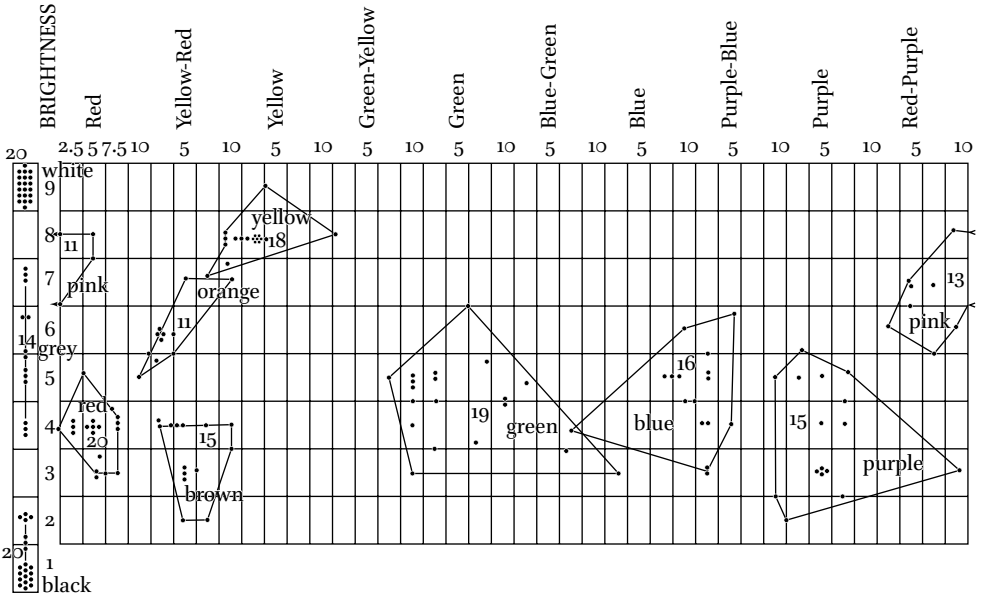

FIGURE 5.1 Crosslinguistic clusters of focal colors
BERLIN AND KAY 1699: 9. REPRINTED WITH PERMISSION

regardless of where it occurs, regardless of the type of referent, rather than, let's say, in Conklin's famous example—so he has this term, *latuy* 'light green and mixtures of green, yellow, and light brown', but this this is only used to label certain kinds of plants. So obviously what you have here is a conflation of color and some sort of stage of a plant in its life cycle. So these terms are not basic color terms in Berlin and Kay's sense.

Given the basic color terms of each language as identified by this set of criteria, we can ask, what are the best possible instances? And that's what Berlin and Kay did. The result is that the best instances, the focal instances, fall into these 11 clusters. On the other hand, we can ask if it then is the case that what languages lexicalize primarily are those focal colors. Because Berlin and Kay found much more variation in the extension of the categories than they did in the focal colors. So the boundaries of the extension would vary a great deal not just across languages, but also across the speakers of any one language. And of course this is all part of what gave rise to the idea that categories, such as color categories, are cognitively represented in terms of a prototype and some sort of principle according to which the other members of the extension are related to the prototype, such as some kind of similarity.

So if it is in other words primarily the focal colors that are lexicalized across the languages of the world, then Berlin and Kay also found that the consistency of the set of focal colors that is lexicalized in a given language is not random. Rather there are implicational generalizations along the lines of a language cannot have a term for red unless it also has terms for black and white, cannot have a term for green and/or yellow unless it also has a term for red, and so on and so forth. Berlin and Kay interpreted this set of implicational generalizations to the effect of an evolutionary scale, going with the idea that was originally proposed by Hugo Magnus a hundred years earlier.

So the new system starts out, not with white and black, but rather with a category that comprises all light and warm colors and one that comprises all dark and cool colors, and then these are in subsequent stages broken down until you get to a level where all the categories that are lexicalized are monofocal, and from there you're supposed to be able to add additional categories.

We talked on Monday briefly about Kay and McDaniel's attempt at coming up with a neurophysiological explanation of the Berlin and Kay generalizations. Also in the 1970s, Berlin and Kay and colleagues conducted the World Color Survey, which involves the collection of 110 color term systems, and the analysis of that went on until very recently—the comprehensive data was finally published in 2009 by Kay and Luisa Maffi and colleagues.

I already mentioned that after much debate, it has been generally accepted that there are languages without basic color terms, and indeed you know

Gladstone's point—languages such as Latin and ancient Greek actually are among them, which was pointed out in the 1990s by Sir John Lyons. Steve Levinson showed pretty conclusively in 2000 that languages don't have to have basic color terms using Berlin and Kay's methodology. Obviously that's hard to do with speakers of Latin, since there aren't any speakers of Latin around.

Obviously Berlin and Kay's generalizations have generally been taken as a strong confirmation for a biological cognitive basis in linguistic semantics which is non-language-specific, and which suggests that rather than language influencing internal cognition, it's really very much the other way around. There is a basis to cognitive conceptualizations, which in its turn forms a substrate to the categories expressed in different languages.

That view has more recently come under a lot of fire, in particular from Debby Roberson and colleagues, who have done several studies that have found evidence calling into question some of the assumptions that were made between Berlin and Kay's and Eleanor Rosch's work.

So in particular, Rosch famously studied the Dani language of Papua New Guinea in the early '70s as part in fact of the original project of Berlin and Kay's. And she found that speakers of Dani, even though this was supposed to be a Stage-1 system, with just two basic color terms, that their memory for all those focal colors that aren't actually lexicalized in there is still better than their memory for non-focal colors. So in other words, color memory would be driven by the same focal categories across speakers of English and Dani regardless of whether they are lexicalized or not. That was Rosch's claim, and it has formed a very important part of how people have interpreted Berlin and Kay's findings.

So Roberson and colleagues looked at a couple of languages that do not have basic color terms. And they found that in fact whatever linguistic expressions for colors there are does very much affect memory for color. So they built a case on the basis of this that suggests that in fact nonlinguistic color categories are strongly influenced by lexicalization, by linguistic expression.

In response, Paul Kay and Terry Regier and colleagues did a reanalysis of the data from the World Color Survey. This is a sort of new take. It leads away from the original Berlin and Kay's hypothesis but suggests instead a somewhat more complex and weaker hypothesis. And part of the underlying assumptions here is that the actual perceptual color space—which they are still assuming as universal—is not what is represented by the Munsell chart, but is something more complex for which people recently have found empirical evidence. A much more complex perceptual color space that is not as uniform as these two equidistantly-spaced dimensions of hue and brightness in the Munsell color chart suggest. It's a space that has "bumps," as Regier and colleagues like to

put it, and those bumps may in fact be driving category boundaries across languages in lexicalization.

Regier and colleagues found some statistical evidence suggesting that against this apparent complex universal perceptual color space, the color term systems of the World Color Survey make what they call near-optimal cuts, meaning cuts that are designed so as to maximize within-category similarity and minimize across-category similarity.

It's not clear that this is in any way inconsistent with what Roberson and colleagues observe. So at the moment, things are in flux. There are a lot of different currents. It's once again a field of a lot of interest and debate, and we're going to have to see where things are going to go in the next few years, but I'm sure there will be quite a bit more exciting research.

The last domain that I want to talk about in terms of the three classic domains of ethnosemantics is the domain of ethnobiology, meaning the classification of plants and animals in different languages. Again the question is, what's the basis for nature and what's the basis for nurture? The basis for nature obviously would be that, you know, we all sharing a planet where the biological species that we encounter are roughly similar, and we share a planet with these species in ways that may be recurrent across cultures. At the same time, we know that there is variation first of all in terms of what species may occur in a particular place. There is moreover variation in terms of the cultural significance of different species, in that their economical use depends on the modes of subsistence and production prevalent in a given culture. Their ritual and medicinal significance depends a lot on the spiritual and physiological theories that are prevalent in that particular culture and so on and so forth. Again we have a basis for possible uniformity. We have a basis for culture-specificity. Let's see how the two pan out.

Again there was also a lot of earlier work in ethnobotany that was primarily out to discover evidence for language- and culture-specificity and describe that. And the seminal study that is nowadays considered to have put ethnobiology on the map in the cognitive sciences was a study conducted by Brent Berlin and colleagues in Tenejapa, in Chiapas, a Tzeltal Mayan speaking community, on the classification of plants in the local environment in the county of Tenejapa.

Berlin brought together a team of botanists, and linguists, and anthropologists, and what these people did is they went botanizing with native speakers, members of the local community. They actually put a geographic grid on a map of the area and went through each cell of that grid and systematically covered the entire area, collecting specimens of all the plants, and had the speakers who went with them label the plants in the wild where they encountered

them. So they collected both specimens of the plants and labels for these specimens. And then they did analyses on the labels with the help of the native speakers of the language.

So in particular they used linguistic tests to identify hypernyms and hyponyms, where a hypernym is a superordinate term and a hyponym is a subordinate term. And they used classification techniques to deal with covert categories, meaning categories that are considered to have a common superordinate category, so they are in a sense co-hyponyms, except that the superordinate category happens not to be lexicalized. How do you discover such a category? Basically you run a similarity test on the linguistic labels, which they did.

So the generalizations that Berlin and Breedlove and Raven came up with in 1974 then subsequently underwent many revisions and were tested against the work of other researchers in other parts of the world. In 1992, Brent Berlin published a book on the results of ethnobiological research of several decades from his perspective. And in this book, he proposed a number of generalizations that he argued are true for ethnobiological classifications around the world.

The first of these is that all ethnobiological classifications distinguish among no more than four to six taxonomic ranks. And as a matter of fact, more recently, it's been pointed out by scholars such as Scott Atran that generally ethnobiological classifications are much more flat than that. So they generally mostly only distinguish between two or three ranks rather than as many as six.

But in any case there're no more than six, generally speaking, which Berlin calls the 'unique beginner or the kingdom level,' e.g., something like *plant* or *animal*; the 'life form level,' e.g., *tree*, *bird*, *fish*; the 'intermediate level' if there's one, such as *leaf bearing tree* versus *needle bearing tree*; the 'genus,' e.g., *oak*, *maple*, etc.; the 'species,' e.g., *sugar maple*, *white oak*; and the 'varieties' such as *cutleaf staghorn sumac*. This is by the way taken from George Lakoff's discussion.

Next, the really interesting question is of course what are the principles according to which these categories are organized across languages. And here Berlin's famous claim is that the principles are in fact the same that are used in the Western scientific classification of plants and animals, the so-called Linnaean classification, which was developed by the Swedish biologist Carl Linnaeus in the 18th century.

This claim has been very controversial, because it [has been understood to entail] that culture-specific principles such as cultural utility play no role in the classification, which actually is not what Berlin has been saying. And the claim also has been misunderstood to the effect that Berlin is suggesting an

innate basis for ethnobiological classifications, which as far as I can tell is not true. What Berlin is saying is that there are more general innate principles of categorization, and when applied to the biological domain, these will yield the uniform properties that can be observed across ethnobiological systems. That doesn't mean that there are any ethnobiological categories or even principles or ethnobiological classifications that are *per se* innate.

Berlin did in fact find a lot of evidence for culture-specificity and also acknowledged this. He argued that the greatest amount of culture-specificity is found both at the bottom and at the top of the taxonomic hierarchy. The distinctions at the species and variety level, this is something that is driven a great deal by culture utility apparently. So, for example, if you think about the many, many different dog breeds that we distinguish, and this is all at the species and variety level. That's clearly a reflection of the fact that dogs play a very prominent role in human society. On the other hand, there's also a great deal of variation at the highest level, meaning the life form level, where languages and cultures vary a great deal in terms of the distinctions they make.

It's really, what Berlin is saying more precisely is that there's one level of categorization, which is an intermediate level, a level in the middle of the hierarchy, and at that level the amount of cultural specificity is the smallest. So that level is the genus level. Above the species level, below the life form level. At the general level, you find categories that are to the largest extent, relatively speaking, based on the morphological appearance of the plant, and on what is easily recognizable in terms of the structure of the plant, including its reproductive organs. But what does not matter so much or least at that level—and this is the level that Berlin famously called the *basic* level of ethnobiological categorization—what matters least at the basic level is cultural utility apparently.

You'll also find the greatest amount of prototype organization at the basic level, in the sense that the basic level is the one where people have the easiest time identifying a focal instance and where you can observe the greatest variation between the focal instance and other more marginal members of the category. Whereas the higher level, the life form level and the intermediate level, there is one that is much more abstract. It doesn't have a concrete prototype. And the lower levels, the species and variety levels, are much more uniform in their extension.

The basic level is also the level at which you find the most dense lexicalization across languages. And the idea is basically that there is this economy of categorization such that the process starts at the basic level where you divide the domain on the basis of perceptual appearance and morphological categorization, categorization of the structure of the plant. And then you form

higher and lower ranks of the taxonomy by on the one hand lumping, abstracting across basic-level categories, and on the other hand splitting, subdividing basic-level category into species and variety-level categories.

What's the upshot here? The upshot is, you know, we found evidence that there's no red/non-red color term system, meaning the amount of crosslinguistic variation, the amount of language-specificity in categorization that the structuralists assumed and asserted would be there, would be observable in the languages of the world, was not confirmed. On the contrary, there was found evidence of uniformity in semantic categorization across the languages of the world that could not be explained in linguistic terms.

That's the interesting thing about the absence of a red/non-red color term system. The reason this is such an interesting finding is because there's no linguistic explanation for it. Linguistically, like I said, this is a very plausible system. If it doesn't exist, the explanation for its absence can only come from cognition. And so these findings of apparently universal traits in categorization that manifest themselves in semantic systems of the natural world across languages, the world over, play a very important part in the rise of the cognitive paradigm, had an enormous influence in cognitive science, and it also contributed to the demise of the Linguistic Relativity Hypothesis in its classic manifestation, meaning the ideas that the categories of thought are largely derived from linguistic categories, categories of the native language, which is something that Whorf did apparently believe.

At the same time, we don't know nearly as much about semantic categorization as many cognitive scientists have been asserting since the 1970s under the impact of a popularized image of in particular the results of Berlin and Kay. And in actual fact, the findings, even Berlin and Kay's generalizations, are still compatible with a significant influence of language on internal cognitive categorization. As Paul Kay and William Kempton in fact demonstrated in 1984 when they came up with the first empirical proof by contemporary psychological standards of a Whorfian effect of linguistic semantic categorization onto conceptual internal categorization.

And of course we still know very little about the semantic systems of many, many languages, because we haven't actually looked at many languages upclose, especially when it comes to semantic categorization. There are between six and eight thousand languages spoken on the planet. We have even the most modest scientific records of maybe half of them at most. That's my colleague Matthew Dryer's estimate. I would probably go lower if anything. But when it comes to data on semantic categorization and data that we can use for cross-linguistic comparison, the number is much, much smaller. We have evidence

from a few hundred languages, that's it. So it's still a great deal of possibility for surprises such as the evidence that was uncovered by Debby Roberson and colleagues.

There are many new domains of ethnosemantic research that have opened up or have been opened up over the last ten years or so. Now these are all lines of research that have a more explicitly typological focus, and as I said in the beginning, the people in Whorfian research are mostly no longer anthropologists, but rather linguists and psychologists, so you could just as well present all of this research under the rubrics of semantic typology. But in many ways, it fits the pattern of ethnosemantics in the sense that we're still looking for the categorization of phenomena of the natural world, and we are still interested in cultural specificity.

One domain I want to mention is what some people have called ethnophysiography, which is the ethnosemantics of land and water forms. These are the ethnosemantics of the natural environment, the local geography. The term 'ethnophysiography' was coined by David Mark, who is a professor in the Geography department at my university. A student of mine defended her dissertation a couple of years ago with a study of the ethnophysiography of Seri, which is an unaffiliated language spoken in the state of Sonora on the coast of the Gulf of California, the sea of Cortez, in northern Mexico.

There was also an edited volume of studies in ethnophysiography, a special issue of the journal Language Sciences edited by Nicolas Burenhult and Stephen Levinson in 2008. Carolyn and I added a paper to that one as well.

The people at the Max Planck Institute for Psycholinguistics, my former colleagues, have also been looking at the structure of the human body and how that is expressed across languages, the divisions that different languages impose, the sharpness of these divisions, and so on and so forth. That's very interesting work. And, as part of the project on spatial categorization in Mesoamerican languages which I will be talking about more in the coming days, we've been looking at meronymies, terminologies for body and objects parts in languages of Mesoamerica. We are now looking at languages from around the world.

And again at the Max Planck Institute, Asifa Majid and Gunter Senft and Steve Levinson and colleagues, have been looking at the categorization of tastes and smells more recently, so across languages what categories for taste and smells are lexicalized. Obviously a fascinating topic, and there too is actually early research going back to the late 19th and early 20th century. And then for a long time there was nothing. Now people are looking at this problem again, and Asifa Majid just got a huge grant from the Dutch government

to form her own research group and study this phenomenon some more. So much more research going on again recently in this area, and there should be more fascinating results coming in over the next few years to be sure.

I believe there's indeed robust evidence for traits that are shared across many different languages, many different semantic categorization systems, and those traits we need to explain. I still believe we can explain those with reference to principles of categorization. It doesn't have to be innate categories, as I said on Monday. Maybe—and in fact this is something that was foreshadowed in Brent Berlin's work on ethnobiology—what's innate and shared across speakers of different languages and members of different cultures are just the underlying principles, strategies of categorization, not the categories themselves. And that in turn gives the input from language a much broader potential role in forming culture-specific categories, in informing culture specific categories, and effectively serving as a means of transmission of culture and knowledge between generations.

LECTURE 6

Semantic Typology: the Crosslinguistic Study of Semantic Categorization

Having concluded the first half of the lectures, we're finally getting to my narrow area of specialization: semantic typology. And this is going to be the introduction to it.

So I'm going to give you the short answer to the question, what is semantic typology? Then we're going to look at linguistic typology in general, for some background. I'm going to introduce the method for doing semantic typology that, well, my former colleagues at the Max Planck Institute for Psycholinguistics tend to call it the 'Nijmegen method,' although people outside the Max Planck Institute could take issue with that label, because it's not obvious that the Max Planck Institute can lay claim to being the sole originator of this method, or even historically the first originator.

I mentioned previously the issue and the problem of etic grids, that's something we will need to talk about, and that pretty much sums it up for today.

For my introduction to the problem of semantic categorization, Let's take an example from the film *Wall-E*. So a little robot is sorting his prized possessions, and he's coming across a problem. He is trying to classify an implement that Americans call a *spork*, which is combination of a spoon and a fork. He has a category of spoons, and he has a category of forks, and this one falls right in between, which is the conclusion that he comes up with eventually.

What the heck does that have to do with semantic typology? Nothing in particular and a lot depending on how you look at it. What this illustrates are the basic properties of categorization more generally. We have categories that we try to subsume stuff under. We like to sort it into one bin or another. What's interesting is we use these categories for thinking about stuff, for identifying it, for identifying the properties, and so on and so forth.

All original audio-recordings and other supplementary material, such as any hand-outs and powerpoint presentations for the lecture series, have been made available online and are referenced via unique DOI numbers on the website www.figshare.com. They may be accessed via this QR code and the following dynamic link: https://doi.org/10.6084/m9.figshare.11419131

© JÜRGEN BOHNEMEYER, 2021 | DOI:10.1163/9789004362628_007

But we also use categories when we talk about stuff. Linguistic reference to things in the world—objects, events, actions, states, and what have you—require a classification of these entities or states of affairs, and it requires a classification in terms that are prescribed by the conventions of the language you're using. That's the problem of semantic categorization. And the study of that is basically semantic typology.

We also have the problem of sorting colors into semantic categories, which we've already discussed. There are languages that have basic color terms for green and blue, such as English; there are languages that have a single basic color term for green and blue, but not for other colors, such as Yucatec; and then there are languages that have basic color term for green, dark blue, and light blue, such as Russian and Turkish.

Semantic typologists study such semantic categorization systems in individual languages and then they try to describe the distribution of these properties across the languages of the world. As is represented here by this map from the World Atlas of Language Structures, which gives you the distribution of languages that have a single category for green and blue, versus languages that distinguish at the level of basic color terms between green and blue.

Once we've mapped out the distribution of semantic categories across the languages of the world, we try to formulate implicational generalizations over them. The best-known example is Berlin and Kay's evolutionary scale over possible basic color term systems, which we discussed this morning.

Let me talk a little bit about the history of this enterprise. It really started in the 19th century, in the late 19th century with studies such as Louis Henry Morgan's kinship term questionnaire, which we talked about this morning. Charles Darwin was actually an early pioneer of semantic typology, believe it or not. He did also a questionnaire study on gestures, gestures in different languages and cultures, co-speech gestures and their meanings. We also already talked about Hugo Magnus' study of color terminologies, which was essentially based on a methodology similar to Berlin and Kay's. He used a smaller stimulus kit, but a larger sample of languages.

Now, I also mentioned this morning already that this line of research pretty much went dark in the beginning of the 20th century because much of it was marred by invalid assumptions—racist and social Darwinist ideology and so on—but also by a lack of sophisticated methods of linguistic semantic analysis.

So, that's where, you know, we perhaps have come a little farther in the present and therefore have picked up the thread once again, and resumed the enterprise of semantic typology, and maybe made some progress, maybe. So in between, however, there was a second phase, and that was Ethnosemantics, an enterprise primarily seeded in cultural and linguistic anthropology, and

that's what we talked about this morning. So that was research that did look at semantic categorization, but not so much from a typological perspective, but rather primarily in terms of studies that would look at one language and one culture only.

This research program was very much informed by the Linguistic Relativity Hypothesis, by attempts at confirming or disconfirming the Linguistic Relativity Hypothesis, or by attempts at doing what Whorf actually suggested to the scientific community should be doing, which is not to test the Linguistic Relativity Hypothesis, because he never thought about a Linguistic Relativity Hypothesis. The research program that he suggested was one of studying cultures in the terms prescribed by the native language of the speakers, of the members of the culture. In other words, what Whorf called the Linguistic Relativity Principle was really all about describing cultures in their own terms, something quite different from how it came to be interpreted later.

So then, the third and contemporary phase is the resurgence of semantic typology, which is actually part and parcel of the resurgence of typology in general. Typology, linguistic typology, syntactic, phonological typology, has seen a history similar to that of semantic typology. It was originally a 19th century enterprise. Then it pretty much became dormant, or as people no doubt thought at the time, extinct, until it was revived by Joseph Greenberg in the early 1960s.

In the 1980s, there were a number of large-scale studies on semantic categorization across languages, which however used primarily questionnaire data. Dahl's study on tense-mood-aspect systems used questionnaire data—we actually talked about Dahl's questionnaire yesterday, whereas Åke Viberg's classical study on perception verbs, on semantics of perception verbs across languages, and of course Talmy's of lexicalization patterns and motion descriptions, these two were not primarily based on original data collected by the researchers but rather on secondary sources, on available descriptions of the relevant languages.

There are other recent studies, some of which we'll still discuss as we go along, such as the one by Pederson and colleagues on spatial frames of reference and spatial categorization in 13 languages. In 2003, Stephen Levinson and Sérgio Meira and the Language and Cognition group, of which I was a member then, did a study on the semantic similarity of topological spatial relations across different languages, focusing specifically on adpositions—prepositions and postpositions—on the basis of data collected with the BowPed picture series, which has come up already various times in the lectures. And more recently, Khetarpal did another study using a different style of analysis of the same set of data.

I did a project with two colleagues in 2006 on motion event categorization based on Len Talmy's typology, and here the question was to what extent does this influence the nonlinguistic categorization of motion events? We looked back then at a sample of 17 languages, but we're still working on this. And we actually are testing speakers of additional languages as we speak now.

I also already mentioned the study on verbs of cutting and breaking that we did at the Max Planck Institute in the beginning of the last decade. I did a paper on the argument structure of verbs of cutting and breaking, based on data from 17 languages again. This is coincidence, it's not the same 17 languages. Meanwhile, Asifa Majid, Melissa Bowerman, and James Boster did a paper on the basis of the same sample, well, the same data set, but from a larger number of languages, a study on the semantics of verbs of cutting and breaking. Terry Regier and colleagues did an analysis of the semantic similarity of color terms in the languages of the World Color Survey.

So there is a lot of ongoing research. It's a very active field right now, and you can see a trend: the corpora, the samples of languages that we are working with slowly get larger, and there's a drift towards statistical methods. Slowly but surely statistics and quantitative methods are taking over, which is the reason why we will be talking about quantitative methods in the second lecture tomorrow.

As a quick reminder, by way of background, there are background assumptions that have been influencing this research project from the get-go. Before the cognitive turn in the 1940s and 1950s, it was widely assumed by linguists and anthropologists and psychologists that a large part of thinking proceeds in terms that are modeled after the semantic categories of the native language. These categories in turn would be culture-specific. Then came the cognitive turn, which put cognition as a pre-structured phenomenon, something that has some sort of information that the child is born with from the get-go, that's not a blank slate. And that has led over time to a very strong emphasis on innateness, which assumes that large parts of cognition are governed by innate processes and principles, and language and culture really only play a fairly shallow and superficial role in cognition.

Finally you have the Neo-Whorfian program starting somewhere in the late 1980s probably, with an attempt at putting culture back into its rightful place in the theory of cognition. At least from my point of view, that's basically the name of the game. That's what we are trying to do.

That's semantic typology in a nutshell. In order to get a better understanding of how this works, how we are getting the data that we need, and how we'll analyze them, we have to understand a little more about linguistic typology in general. So let me talk a little bit about that.

What is typology? Typology is the inductive or bottom-up approach to language universals. That's one way of putting it. You don't have to have the aim of finding language universals when you are doing linguistic typology. But what matters is you are going to come up with generalizations over the distribution of properties across the languages of the world, and these generalizations are therefore valid for the total number of extant languages. In that sense, you could say they are universals are typically not exceptionless, absolute universals, i.e., universals along the lines of 'All languages have X or Y or Z,' You will often find statements such as 'All languages have nouns and verbs,' 'All languages have syllables,' 'All languages have consonants and vowels,' and so on and so forth. The fact of the matter is you can almost always find exceptions to these kinds of claims. Instead there will be generalizations of the form 'If a language has property X, then it also has property Y,' such as we saw this morning with respect to the example of Berlin and Kay's evolutionary scale.

So how do you do this? You start out surveying the variation in some domain that you are interested in, which could be a syntactic or morphological domain or a domain of phonological research or a semantic one. Any property of language counts. In fact, the British sociolinguist Peter Trudgill just published a very interesting book called *Sociolinguistic Typology*. Although I think the title is a little bit misleading, because he's not so much talking about a typology of the distribution of sociolinguistic types of languages. He's rather talking about sociolinguistic conditions that favor or inhibit language change. Nevertheless, a very interesting book.

Anyhow, so pick a domain, start out from an initial survey of the variation, try to come up with a list of the discrete types of the phenomena for this domain. From there try to extrapolate a finite list of the possible types, and then try to explain why possible types that don't occur don't occur, and why a given particular language has the type it does rather than another type. This is generally going to lead you to implicational universals, implicational generalizations.

The paradigm for this type of research was Greenberg's classic study of word order universals. I hope I'm not boring you guys to death with this stuff, since I'm sure that people are very familiar with this. But it's important to understand that this was a radical break with the tradition in linguistic typology, the 19th century tradition which did not ask these kinds of questions, like you know, what types of word orders occur in the languages of the world? And how can we formulate useful generalizations about the word order phenomenon that occur in the languages of the world? Rather, the 19th century was primarily interested in classifying entire languages, and trying to explain why languages look the way they do with respect to assumptions about the culture of the speakers, along the lines of whenever a phenomenon looks simpler than

the phenomenon in other languages, that was assumed to reflect a more primitive culture, or something like that.

Let's talk about Greenberg's typology of word order universals very briefly. So assume a typology of the order of the main constituents in transitive clauses, and assume it is possible to somehow identify the arguments of the transitive verb as a subject and an object. However you do that, which you know it's really very unclear because it's not very clear that all languages have subjects and objects. But, you know, on purely semantic grounds, so maybe we should call the subject an agent, or an actor, and the object a patient or an undergoer, something like that.

On some such grounds, let's say we can identify the syntactic arguments of the transitive verb. So then among languages that have a single preferred order, we can come up with six logically possible types SOV, SVO, OSV, OVS, VSO, and VOS, and aside from these, there are languages that have multiple types. There are also languages, for example, my own native language German, has one type in main clauses and another type in subordinate clauses. There are also languages whose word order is so free that it is impossible to establish a type like this.

But, if you look at the distribution of these types across the languages of the world, you'll find that two types, namely, SOV and SVO, are extremely frequent, SOV of course being the type of, for example Turkish and Japanese, and SVO that of Mandarin and English. Then, a much less frequent type is verb-initial and has the subject either preceding the object, as in Welsh or the Mayan language Jakaltec, or even less frequently a language that has the verb in initial position and the subject in final position. That happens to be the case in Yucatec, the language that I work on, but also for example in Malagasy. Object initial languages are really exotic. The type that has the object in initial position and the verb in final position is apparently still unattested.

If you establish these types, you can start formulating implicational generalizations. So for example, Greenberg proposed that whether a language has prepositions or postpositions correlates with whether the language puts the objects before or after the verb. So verb-object languages tend to have prepositions, whereas object-verb languages tend to have postpositions.

When I say "tend to have," you get the sense that I'm talking about quantitative generalizations, in that you can use statistical methods to test these, which in turn entails that you are going to benefit from having large samples of languages you can look at.

Similarly, verb-object languages tend to put the possessor after the head in nominal possessive constructions, so they tend to say something like *the ball of Peter*, whereas object-verb languages tend to have the opposite order, so they

tend to say something like *Peter's ball*. Of course there are alternative languages that, like English, have both orders. Greenberg also proposed that whether a language has attributive adjectives preceding or following the nominal head correlates with whether they put the object before or after the verb.

So more recently, my colleague Matthew Dryer compiled a very large database, which now has over a thousand entries, meaning it has information for over a thousand languages, which now contains information on around 1,800 languages. He compiled this database from existing language descriptions. In this respect, the syntactic and phonological typologists typically have an edge over the semantic topologists, because they can rely on existing language descriptions in a way that the semantic typologists typically cannot. Available descriptions, grammars, and dictionaries of the languages of the world typically don't give you much information about the semantics of these languages.

Matthew compiled a database, and against this database performed statistical tests of Greenberg's proposed universals, and confirmed most of them, but disconfirmed some. So the order of the attributive adjective with respect to the noun head does not actually statistically correlate, not significantly anyways, with verb-object versus object-verb order.

So then once you've found a set of generalizations, you can try to explain these generalizations. A number of explanations have been proposed. One often finds informal explanations in terms of grammaticalization in the literature. More formal statements include Mark Baker's 'Head Directionality Parameter.' As the name 'parameter' already suggests, this is basically within mainstream Generative Grammar.

Dryer himself developed what he calls 'Branching Direction Theory,' which basically says a language tends to be consistently left-branching or right-branching, and that in turn can be explained in terms of processing. So the idea is that if a language has consistent left branching, or consistent right branching, that might ease processing relative to a language in which either order occurs with roughly even frequency.

People have tried to formulate the generalizations in Optimality Theory. Most recently an article by Michael Dunn and colleagues appeared in the journal *Nature* last year, in which they tried to show that word order patterns actually vary much more across language families than they do within language families, which they attribute to extant word order patterns basically being properties of linguistic lineages rather than being governed by universal principles, be it principles of universal grammar, as Mark Baker claims, or be it universal principles of cognition, as Matthew Dryer and Jack Hawkins have claimed. This triggered a huge uproar in the field of typology. There was a special issue of the Journal of Linguistic Typology dedicated to this controversy,

which basically had a lot of people coming up and pointing out weaknesses in Dunn et al.'s statistical analysis, but above all pointing out that actually—the message they are making, here, the point they are making has two aspects to it.

One of Dunn et al.'s two implicit claims is that genealogical relationships, and also language contact, are much stronger factors than whatever universal cognitive tendencies, such as processing pressures, might have an influence on word orders. But that is something that typologists have always taken for granted. That is something that has long been known. It has long been known that word order patterns tend to be more consistent within language families, and therefore presumably languages to some extent inherit them from their ancestors.

That does not however mean—and that's Dunn et al.'s second point, and that's where they are going too far—that there are no universal tendencies. However weak they are, they exist. They are manifest even in Dunn et al.'s data, and as a result, they still call for an explanation, and the cognitive explanation, the explanation in terms of processing pressure proposed by Dryer and Hawkins, continues to be valid. At least, once you've taken into account the balance of evidence, it's not challenged by the data presented by Dunn and colleagues.

I mentioned earlier the database that Matthew Dryer compiled. Matthew Dryer and Joan Bybee are pretty much the pioneers of doing statistics on large samples of languages, starting in the 1980s. And the database that Dryer compiled for his study of word order universals, word order patterns in the languages of the world has become the core of what's now the World Atlas of Language Structures, which is a project that hundreds of linguists from all over the world have contributed to.

There's a published hardcover version, but there's also an online version that you can consult for free. There's a list of currently around 140 features, most syntactic, some phonological, some in fact semantic, such as the distinction of color terms systems that we looked at before, which you can click on, and it will give you a map of the distribution of these features across the languages of the world. Moreover, you can combine multiple features, and the search interface will give you an output that shows how the distribution of one feature correlates with that of another feature.

So let's turn to the so-called 'Nijmegen Method.' I have to say I probably should stop using this term, because I've gotten so much flack for using it. I mean I'm not the one who introduced it. Stephen Levinson introduced it. But the problem is once you understand what the method is in a nutshell—if anything, it ought to be called the 'Berkeley Method'—because the blueprint for this so-called 'Nijmegen Method' was really Berlin and Kay's study.

So the core elements of the Nijmegen Method are the etic grid; the set of nonverbal stimuli; the gathering of descriptions of the stimuli from speakers of a large range of independent languages. "Independent" means unrelated, not in contact with one another, and if possible also from different areas. So we're trying to get the best spread possible. Once you get the data, you're going to perform a semantic analysis and a distributional analysis, trying to uncover underlying implicational generalizations along the lines of Greenberg's paradigm.

So let's do this in a little bit more detail. So you start from the etic grid, but actually you start with constructing the etic grid itself. And, so 'etic grid' is a term that has come up several times already. Let me at least now finally tell you why it's called the 'etic grid.' It comes from 'phonetic,' which is odd. Why would a research tool for semantic typology come from phonetics? That's why we get rid of the *phon-* part. So what the heck am I talking about? The idea of an etic grid was originally modeled after something like the International Phonetic Alphabet, which is a phonetic classification of the possible speech sounds that occur in the languages of the world in terms of their articulatory and acoustic properties.

So this is a phonetic classification, therefore it's not a language-specific classification. It does not take into account the function that a sound plays in a particular language, whether it's a phoneme, a meaning-distinguishing element in that particular language, or whether it's only an allophone of some other sound category. So that means phonetic categories are not language-specific, whereas phonemic or phonological categories are language-specific.

So Pike took this opposition between phonetic and phonemic categories, where the phonetic categories are non-language-specific, and the phonemic categories are language-specific, and he generalized it to any kind of contrast between a language- or culture-independent classification and a language- or culture-specific classification.

So an etic grid is a categorization that is language-independent, and it's supposed to hold as a classification matrix for the domain of study. A classification matrix that is valid for all languages and therefore can be used as a frame of reference to map the different category boundaries in the different languages that you're going to study. People have pointed out that this is a very problematic approach, in the sense that it necessarily introduces a bias, and possibly even the risk of circularity. We'll talk about that after our break, which is coming up shortly. It introduces the risk of circularity, which some people have claimed renders studies of this type, studies that use this methodology, invalid. But they haven't proposed an alternative.

So in effect what these people are saying, and when I say these people I mean, for example John Lucy, who is actually somebody who has been

generally highly sympathetic to the research of the Nijmegen group, Stephen Levinson's group. But based on his criticism of etic grids, you have to conclude that he assumes that semantic typology, to the extent that it is based on this method, is itself invalid. So instead of going out and studying a large sample of languages typologically in this manner, we should probably just resort to studying individual languages only. So I don't agree with that obviously. I will tell you what I think about the etic grid controversy and how to deal with etic grids after the break.

So we've seen a bunch of examples of etic grids over the time. In Berlin and Kay's study, the etic grid was the Munsell color chart. That's probably the most famous etic grid in any study of this type. In the classical work on kinship terminologies starting with Louis Henry Morgan, the etic grid is the genealogical network of relationships. In the research on ethnobiology by Berlin and colleagues, it's the Linnaean classification, or scientific classification, of plants and animals, and so on and so forth.

So the etic grid of course doesn't come out of the blue. When you're starting on a project of semantic typology, there's always already some literature that's there, at least maybe descriptions that are based on data from one language only, so those descriptions you can go by when constructing the etic grid. But once you've arrived at an etic grid, which is basically your prediction of what the independent conceptual dimensions of the domain of study are going to be like, meaning the conceptual dimensions along which the categories of individual languages and categories across languages might contrast. Once you've predicted these dimensions, once you've constructed your etic grid, you then proceed to devise sets of predominately nonverbal stimuli that encode the various cells of this grid exhaustively.

It doesn't have to be a set of nonlinguistic stimuli. You can also construct a questionnaire, which is a linguistic approach, which is for example what Östen Dahl and colleagues did for that study. That was also based on an etic grid, although they didn't make that etic grid very explicit. But in Berlin and Kay's study, obviously the cells were encoded as color chips, in other words, nonverbally, for very obvious reasons, because it's very difficult to do a questionnaire study on color terminology. And that's been the predominant approach in the Nijmegen tradition, as well as being the predominant approach in my own work.

You then proceed to collect descriptions of these stimuli under a particular protocol, which has to be fixed. Remember yesterday we talked about the problem of giving the participants a specific task of what they're supposed to be doing with nonverbal stimuli. Remember the example of the BowPed pictures.

The speakers are supposed to locate one object with respect to another, but in order to do that in the sense that is intended by the design of the study, you have to give them an elaborate context in which they do this. So you have to be very careful about how you formulate the task.

But with that protocol, which has to be realized in the same way by every collaborating researcher with every population that is tested, you then go out and collect data from a range of languages as varied as possible. I already mentioned, it's supposed to be genealogically independent in terms of contact, as widely distributed geographically as possible, and I should also say they're also supposed to be typologically independent in the sense of, if you assume a classification of the languages of the world in terms of basic typological parameters, such as word order, or the typology of head-marking versus dependent-marking languages, you want languages that are as different from one and another as possible, so as to ensure that your findings aren't an artifact of some linguistic type and are really valid for that particular linguistic type rather than for all languages of the world.

And of course, as I already explained, you have to do this collaboratively, because it's not going to be possible for any decent-sized sample of languages that you alone can collect all the data. You may be able to collect the data, but you won't be able to analyze it, and you won't be able to have the kind of expert knowledge of all of these languages that's needed to perform any, however modest, level of semantic analysis.

This need to go collaborative, to have a team of researchers that collect the data, has been limiting the size of the language samples, which has been lagging far behind that which is now the standard in syntactic typology. But as we already talked about, the main reason for this difference really is the availability of existing descriptions for the kind of data that you can do syntactic typology on, but not for the kind of data that semantic typologists want to work with and are interested in.

In reality the largest study in semantic typology ever carried out was not actually the World Color Survey, which looked at 110 languages, but the Mesoamerican Color Term Survey, which was conducted by Robert MacLaury and collaborators, which looked at 116 languages of a particular geographic area actually.

From there the number goes rapidly down. In the Nijmegen paradigm, the largest study that we've done was in fact the study of verbs of 'cutting' and 'breaking.' So data for that project has been collected from around 30 different languages spoken all over the world. Now 30, by the standards of the kind of sample that's now pretty much the norm in syntactic typology is a pathetic

number. That's very, very small. But it's the best we can do at the moment in semantic typology.

You have to ideally take into account the possibility of not only inter-speaker variation, meaning variation across different speakers of the same language, but also intra-speaker variation, which means one and the same speaker may give you a range of different responses. People have been using different approaches to this problem. A very popular one is to just simply go with the first response. I'm not actually a huge fan of that. I think it's important to get the range of possible responses and within this range identify the speaker's preferred response. But that of course takes a lot more time and effort.

There is always the possibility of elicitation artifacts. There is always the possibility of limited ecological validity, meaning you're collecting the data on the basis of a task or a set of stimuli that really have no place in the everyday culture of the people you are working with, and therefore could be argued to really not represent the cognitive and linguistic practices of this community very well, meaning the results won't reflect the cognitive preferences and practices of the community.

The way to guard against these kinds of effects is to check the data that you've collected on the basis of your stimulus set, that in turn is based on the etic grid, against other sources of data, and in particular less controlled sources of data. So, to make this a little bit more concrete, let's say you're doing a study on spatial frames of reference, which is something that I've been working on intensively for quite a while. So recently then, aside from controlled elicitation procedures, you also want to look at, for example, local history narratives, descriptions of the local environment or something like that, which are going to produce a lot of situated spatial reference, and you want to see what kinds of frames of reference people use for that kind of stuff.

So you then try to perform an analysis, semantic and pragmatic analyses of the collected extensional data in order to try to get at the underlying senses. That's what we were talking about yesterday. You want to use statistical techniques to analyze correlations. That's what we will talk about tomorrow a little bit. Formulate implicational generalizations, and try to come up with possible cognitive motivations or cultural motivations for these generalizations.

So, why do we do this? I already mentioned the principal goals of this line of research on Monday. One goal is to try to demarcate the boundary between biology and culture in cognition, meaning identify the role culture and biology play in human cognition, and figure out how these two interact.

You're in the process of being able to verify or reject innateness claims, which used to be incredibly pervasive and in some corners of the cognitive

sciences still are, but they are being beaten back now under what I have been referring to as the empiricist turn in the cognitive sciences. Meanwhile, you are also going to be able to shed more light on not just the role language plays in cognition and vice versa, the role cognition plays in language, but also on the principles that govern the mapping between linguistic form and meaning at the syntax-semantics interface. Starting of course from principles of lexicalization, what principles govern what conceptual categories, which conceptual categories get expressed across languages; but ranging all the way to the syntactic side of the syntax-semantics interface, such as principles of argument structure, such as in the study on the argument structure of verbs of cutting and breaking which I eluded to, or for that matter, principles of event segmentation, which we will talk about on Friday.

There are a number of other sources of evidence concerning the nature and workings of the interface between language and nonlinguistic cognition. And ultimately, it's always going to be very fruitful to try to combine the results from semantic typology with these other sources of evidence, in particular evidence from language acquisition research and developmental psychology, evidence from sign language research, co-speech gesture, and of course language processing research. Those are the principal sources that we can draw on when we are trying to find out what cognitive processes are involved when people speak, or prepare to speak, or when they comprehend speech.

Let's talk about the knotty issue of etic grids. So I already mentioned the International Phonetic Alphabet as a sort of model etic grid. Of course it's not a model of a semantic etic grid—it's an etic grid for sound categories. But it is an etic grid nonetheless, because it's a language-independent classification. In this case it's a vowel chart, so this is a language-independent classification of vowel-like speech sounds that occur in the languages of the world, based simply on articulatory properties. It doesn't talk about the role that a particular vowel may play in the phonology of any given individual language. That's an 'emic' question.

So we already went over these examples some more in the more recent research on taste terms, linguistic categories for tastes. Researchers have used a stimulus kit that is based on the traditional four taste senses of European culture—sweet, salty, bitter, and sour. Plus a fifth one, which is *umami*, the category that's associated with monosodium glutamate. Researchers have used samples of foods or chemical substances that are supposed to induce just these sensations and have studied their classifications across languages.

So I've been referring to etic grids themselves as frames of reference, and what I mean by that is illustrated by the Munsell color chart, which you can use

as the backdrop for locating the various language-specific color terms in terms of their extensions, which you can represent as Venn diagrams against the grid of the Munsell color chart. Or if you think of Bowerman and Pederson's Topological Relation Picture Series—i.e., "BowPed"—we saw those Venn diagrams that represent the cuts different languages make against some assemblage of the BowPed pictures. You can think of that as an etic grid, although that's not the etic grid that was originally underlying the study.

Now, as I said, people have criticized the approach to semantic typology on the basis of etic grids, charging that the use of etic grids will bias the findings of your typological study and possibly make the generalizations circular. So this is what we need to take a closer look at. So in the study of kinship terminologies, it has been charged that the network of genealogical relations that's been used as the etic grid since all the way back to the days of Louis Henry Morgan has been covertly based on assumptions of biological classification in kinship, and these assumptions are not universally valid for all languages. And researchers have argued that we should do away with this methodology, but the problem is, if you get rid of this grid, then how are you going to compare the meanings of different kinship categories across languages?

The problem is for any kind of comparison, you need a *tertium comparationis* as it's called in Latin, you need to standardize the comparison, you need some kind of frame of reference or map, whatever your preferred metaphor is, some kind of conceptual space such that you can show how the content of the semantic categories of different languages relate to one another. That's the problem.

So in the research on color terminologies, John Lucy and Saunders and van Brakel have charged that the findings that Berlin and Kay proposed in 1969 and that, with considerable modifications, are still being upheld by the proponents nowadays. They are essentially circular in the sense that, supposedly one finding of Berlin and Kay's study was that all languages have basic color terms and basic color terms are lexicalizations of the kind of distinctions that are captured in the Munsell color chart. In other words, they are categories that are based on hue and brightness, and that essentially means that in this approach you're starting out assuming that a classification that is valid for English is applicable to other languages, and then you discover that other languages indeed have a classification system that's like that of English.

Consider this quote from John Lucy. He says, "What about the success of the [Berlin and Kay 1969] approach? After all, as apologists of this tradition often note, it works! These color systems are there! Surely that is an interesting and important fact in its own right. Well I agree that something is there, but exactly what? I would argue that *what is there is a view of the world's languages through*

the lens of our own category, namely, a systematic sorting of each language's vocabulary by reference to how, and how well, it matches our own."[1]

In other words, what Lucy is saying is, the Munsell color chart is really an expression of how color is categorized in languages like English, and when we use that as an etic grid, basically we are asking the question of how well other languages conform to these Anglocentric or Eurocentric categorizations. There are some problems with this criticism. In many ways it's overblown. For example, some of it is based on a misunderstanding of their claims. To some extent it's also true that Berlin and Kay's claims overshot their own empirical basis, what in other words the data they found actually supports, and they had to cut back on these claims considerably.

At the heart of it, where the charge succeeds is in the claim that the use of etic grids invariably biases the possible outcomes of your study, and biases them in terms that are informed by the distinctions that you put in the etic grid when you construe it. If you do that on the basis of data from a particular language, then that means the categories of that particular language that you start from are going to bias the outcomes of your study. So that is absolutely true. Lucy and Saunders and van Brakel are absolutely correct about this point. So specifically, when you are collecting data with stimuli that are based on a particular etic grid, then you will not be able with this approach to find any properties of the language-specific categorization of these items that is not captured by the dimensions that already fall into your etic grid. So in the case of the design of Berlin and Kay's study, if there are language-specific categories that have dimensions other than those of hue and brightness, then those dimensions are not going to show up in your responses. If you use the etic grid to map your responses, if you use the etic grid as a frame of reference to compare the responses that you get from speakers of different languages, the effect of conceptual dimensions not represented in the etic grid is not going to show up.

This means that you should never confuse the result that you're obtaining with the stimulus kit on the basis of a particular etic grid, with the categories of the language itself. What you're getting as a result of such a study, such as Berlin and Kay's study of basic color terms, is a projection. You're getting a projection of the categories of the languages that you are looking at onto the etic grid.

At the same time, it's important to understand that this kind of projection can be identified for what it is, so it's not a trap. And a good example of that is

1 Lucy, J. A. (1997) The linguistics of color. In C. L. Hardin and L. Maffi (eds.), *Color Categories in Thought and Language*. P. 331. Cambridge, England: Cambridge University Press (italics in original).

the claim from Berlin and Kay's study that all languages have basic color terms, meaning terms that fulfill all those criteria that Berlin and Kay defined as a threshold for identifying basic color terms. They have to be autochthonous, mono-morphemic, general currency, and so on and so forth, and obviously have to partition some area of the Munsell color space.

So this is in fact a claim that was not sufficiently supported by Berlin and Kay's own data, so they overshot on that one. That was demonstrated in the classic study by Stephen Levinson on the language of Rossel Island a few hundred kilometers off the coast of Papua New Guinea that I already mentioned earlier. But Levinson demonstrated this on the basis of Berlin and Kay's own methodology, using the very same approach, the color chips, the same technique of identifying basic color terms, by Berlin and Kay's criteria, and looking for their extensions on the Munsell color chart, and the result of this study was that this language doesn't have basic color terms. And Levinson was able to demonstrate this in terms such that Berlin and Kay, or Paul Kay, the one who is still pursuing this research on color terms actively today. He accepted this outcome and modified the claim, and it's now no longer a part of the generalization they propose that all languages have basic color terms.

So I would like to propose some guiding maxims for working with etic grids. The first is when you design the etic grid, you should take into account evidence from as large and as varied a set of languages as you can possibly get your hands on. So you should be careful not to just look at English or Mandarin or what have you, one language, and think based on the data from one language, you get an idea of what might be interesting potential dimensions, good candidates of dimensions of crosslinguistic contrasts, and design your etic grid, your possibility space on that basis. You should take into account what is reported in the literature, kinds of variations, phenomena that might drive category distinctions in particular languages. You should take all that information into account when you're designing your etic grid.

Secondly, I already mentioned this—never confuse the projection of a term's meaning on the etic grid with the meaning of the term itself. So you're going to be able to see maybe a term showing up somewhere here encircling an area on a diagram of the Munsell chart, so that may look like it's yellow, so you might be tempted to claim that the term is synonymous with English *yellow*. In fact this term might have meaning components that English *yellow* doesn't have and vice versa. So be mindful of the fact that as long as you're talking just on the basis of data collected with the etic grid, you're not actually talking about the meaning of the terms themselves. You're talking about projections of these meanings.

And then, of course, check the ecological validity of the data against data from other sources of evidence, and in particular from less controlled sources of evidence, ideally from the observation of spontaneously occurring speech events or from evidence gained from participant observation.

Whenever you encounter evidence of aspects of the grid that aren't cross-linguistically valid, you should ideally revise the grid and start over, which of course is often a hard thing to do, especially if you've already established a longstanding research tradition on the basis of what you started out with, as in the case of Berlin and Kay. But it's the right thing to do.

LECTURE 7

Framing Whorf: Reference Frames in Language, Culture, and Cognition

This morning I'm going to talk about a line of research and an intellectual debate that has influenced my own thinking about the relationship between language, culture, and cognition. Today we will discuss the research on spatial frames of reference.

I'll begin with a little bit of background. Let me start with the notion of a place function in Ray Jackendoff's Conceptual Semantics framework. A place function is a mapping from an entity, an object, into a region of space that is defined with respect to it. We can say, using a classification that was originally developed by the developmental psychologists Jean Piaget and Bärbel Inhelder in the 1950s, that place functions in Jackendoff's sense come in two different flavors, namely, what Inhelder and Piaget call 'topological' functions and 'projective' functions.

Typological place functions are orientation-free, perspective-free place functions. What does that mean? If I say *The bandage is on the shin*, that's going to be true, no matter what the orientation of the shin, no matter what the orientation of me, the observer, no matter what the orientation of the configuration in absolute space with respect to the environment. That's a use of 'topological' that is derived from the term as it's used in mathematics, but it is not quite the same sense.

On the other hand, you have projective, or framework-dependent place functions, and those are place-functions that require some sort of conceptual coordinate system to identify that region of space that the entity, which we can identify as a referential 'ground' in Len Talmy's framework, is mapped into. So in other words, what distinguishes the projective from the topological place functions is that the projective ones require some sort of mental coordinate system in order to identify the particular region of space.

All original audio-recordings and other supplementary material, such as any hand-outs and powerpoint presentations for the lecture series, have been made available online and are referenced via unique DOI numbers on the website www.figshare.com. They may be accessed via this QR code and the following dynamic link: https://doi.org/10.6084/m9.figshare.11419134

© JÜRGEN BOHNEMEYER, 2021 | DOI:10.1163/9789004362628_008

What's the coordinate system, then? That's of course our frame of reference. It's a system of axes, which intersect in some point, just like a Cartesian coordinate system in analytical geometry, except the axes are derived from some other entity or feature, and the entity can be the reference entity itself, the ground, in Len Talmy's sense.

So if we want to situate a little toy man with respect to the tree—the toy man being the figure, the tree being the ground—we could try to project a reference frame from the tree alone, that would be an example of an intrinsic tree. With respect to a tree as ground, that's difficult to do, because a tree doesn't have an intrinsic front-back axis. A tree is basically, in the way we tend to think about it, symmetrical all around. Of course that's not really true, but it doesn't have a dedicated front or back.

With respect to a standard chair, we can say that the chair has an intrinsic 'front' and 'back' and 'sides,' and so I can say that a bottle is now 'behind the chair' in an intrinsic sense, meaning in a frame of reference that is projected from the geometry of the chair itself. It can also be 'in front.'

But that's only one way of deriving the axes of the coordinate system. Another way is to project the axes from the body of the observer, in this case, me, the speaker. So I could say that right now from my perspective the bottle is to 'the right of the chair.' That's at least one way of doing it in the way that is supposedly conventional in English.

But in this case, let's say I'm going to put it behind the chair. Most English speakers would say that the bottle is 'behind the chair' in relative terms, since for English speakers the relative front of the chair is the front of the observer projected onto the chair, meaning it's the mirror image of the observer, if you think of the chair as the mirror. And the back, the relative back, or behind, of the chair would be the side opposite the front. Whereas in the West African language Hausa, supposedly it is conventionally the other way around, meaning the 'front' is the side corresponding to the observer if the observer were in the position of the ground but with the observer's orientation preserved. In other words, the region that would correspond to my front region would be this region, so a Hausa speaker supposedly would conventionally say that the bottle is now 'in front of the chair' from their perspective. In my experience, I'm a little bit dubious about this analysis, because in my experience this is a strategy that's available in all sorts of communities—not necessarily in all of them, because I clearly haven't tested all of them—but it's possible for English speakers, although it's a minor strategy for English speakers. It's possible for Yucatec speakers. It's possible for Spanish speakers. They all get it. It's just a minor strategy. So it may very well be the case that there are communities out there for whom this is a majority strategy, as has been claimed for Hausa.

Now this is not the only way of doing it. Another way of projecting a coordinate system onto this reference entity, our chair here, is to derive axes from the environment, and we can do that in concrete terms. So we could say, for example, that the bottle is 'toward the window from the chair.' That would be a so-called landmark-based frame of reference. Or we could say that the bottle is 'toward the door'—same thing. If we had, let's say, a salient river nearby, we could say if the river flows one way, the bottle is now 'upriver from the chair,' and right now it's 'downriver from the chair.'

But we could also derive these axes from the places where the sun sets and rises, and of course, you know, being a habitually egocentric Westerner, I don't have any idea of my geocentric bearings in this room here, but supposing we do, we can now we can say that the bottle is 'east of the chair.'

There are a lot of different classifications of frames of reference. There is, in particular, a difference between the way psychologists customarily classify frames of reference and the classification that has proven most useful for the purpose of linguistic typology. Psychologists tend to distinguish frames of reference in terms of the basis of the axes alone, and the axes are derived from what we call the 'anchor' of the frame of reference.

If I use a relative frame of reference, I project a frame of reference from my body as the observer. That means my body is the anchor of this frame of reference, not the origin—the origin is always some point inside the ground, probably by default the volumetric center of the ground. If I use a riverine frame of reference, an upriver/downriver system, then the anchor would be the river. In a mountain slope system, it's the mountain. In an intrinsic frame of reference, of course, the anchor is the referential ground.

So the psychologists distinguish conventionally between egocentric and allocentric frames of reference, where the egocentric ones are derived from the body of an observer, although psychologists tend to restrict the observer to the cognizer, the person who is actually producing the cognitive representation that involves the frame of reference.

If I say *The bottle is on your left of the chair*, that means it's left of the chair from your perspective. Now this means you are deriving a frame of reference from the body of an observer, but the observer is not the speaker, it's the addressee. In my view this is still an observer-based frame of reference, but the psychologists would say that since it's not your body that serves as the anchor of this frame of reference, this is not an egocentric frame of reference.

Of course that means that the same utterance is egocentric from the point of view of the addressee, but not from the point of view of the speaker, which is absurd. But this is a problem that makes sense to linguists, that linguists

understand as a problem. Psychologists don't understand it. They never dealt with this problem.

I will give you one more. *When you enter the library, the checkout desk is on your left*. Now this has two interpretations: one that works the same way as here, which I get in case *you* is actually meant as an addressee pronoun. But another possibility in English is to use *you* impersonally—at least colloquially, that's possible. In that case, well, who is the observer? It's not the speaker, it's not the addressee, it's a generic observer. It means whoever enters the library, when that person forms a frame of reference as they enter, projected from their body, the checkout desk will be on their left. So this notion of egocentric frames of reference as it's used in the psychological tradition is an over-simplification.

Anyway, so the psychologists distinguish between egocentric and allocentric frames, meaning frames that are derived from the body of an observer versus frames that are derived from some other entities, which can be the referential ground, so that would be an intrinsic frame, or it can be an entity or a feature in the environment, and that would be a geocentric frame.

On the other hand, for the purposes of linguistic typologists, Stephen Levinson proposed a new classification of frames of reference in the 1990s, which a lot of people have misunderstood as a notational variant of the psychological classification, but it's not intended that way. So Levinson's terms are 'relative,' 'intrinsic,' and 'absolute.' Now 'relative' is not the same as 'egocentric.' Not all egocentric frames are relative. Relative frames of reference actually involve a projection of the axes of the coordinate system from the body of the observer onto the referential ground.

But there is another type of frame of reference, which is also based on the observer, but doesn't involve projection. And by 'projection,' I mean in mathematical terms translation. Actually, in the supposedly English-style conventional relative frame of reference, the operation involved here is actually translation plus mirroring. I suppose you could say mirroring alone—reflection in other words, whereas in the Hausa-style realization of the relative frame of reference, we would be dealing with translation.

Now suppose you have a frame of reference that's also based on the body of an observer, but it's not projected; it's not translated; it's not mirrored. The reference entity, and therefore the origin of the coordinate system, is actually the body of the observer itself. So in that case, instead of projecting the coordinate system onto the ground, you're just extending it out into space. You're extending the axes of your own body out into space.

An example of that would be to say *The ball is in front of me*. So this is from a psychological standpoint also an egocentric description, since it's based on the

perspective of an observer. It uses a coordinate system that's based on the body of the observer. But it's not a relative frame of reference, but rather an intrinsic frame of reference in Levinson's classification. Levinson's classification is based not just on the identity of the anchor, but also on the processes involved in deriving the axes of the coordinate system from the anchor.

And why is this useful for typology? This type of frame of reference is rather restricted across the languages of the world. Earlier, I mentioned briefly that this was a really revolutionary discovery. Until the 1970s, it was generally assumed that everybody not only is able to use relative frames of reference, but actually prefers to do so in small-scale space, regardless of language or culture. Then it was discovered that many Australian aboriginal populations prefer geocentric frames of reference even in small-scale space, and that first put the kind of variation that's underlying this entire debate on the map.

So this type of frame of reference turns out to be much more restricted than psychologists and cognitive scientists and linguists used to assume until the 1970s and 1980s. This type of frame of reference, as far as we know, is almost universally available, because there is only one language on the planet for which it has ever been claimed that they use practically no intrinsic frames of reference. All languages and cultures that can use intrinsic frames of reference can also use these kinds of intrinsic frames of reference that use the body of the observer as both anchor and ground, in other words, egocentric intrinsic frames of reference.

Actually, Eve Danziger has proposed the term 'direct' frame of reference for this subclass of intrinsic frames of reference, intrinsic frames that are based on the body of the observer as the reference entity. As far as we know, such 'direct' frames of reference are available to speakers of all languages, with the possible exception of the Australian language Guugu Yimithirr spoken on the Cape York Peninsula, which is that language that John Haviland studied in the late 1970s to early 1980s, which first introduced the idea that this kind of linguistic variation in what frames of reference members of different communities prefer in discourse may also affect the way they think about space. Because Haviland discovered that Guugu Yimithirr speakers, when they talk about historical events, events that happened in real space, will try to situate the iconic gestures that represent the actions and objects in these narrated events in real space, preserving the actual geocentric orientation. So Guugu Yimithirr has been analyzed by Haviland and Levinson as using exclusively geocentric frame of references.

We have within what the psychologists consider allocentric, meaning non-egocentric frames of reference, a family of frames that are based on features of

the environment, meaning the anchor is neither the reference entity nor the observer, and that's what the psychologists call geocentric systems.

Levinson restricts his class of absolute frames to a subtype of geocentric frames that involves not merely projection, meaning translation and/or mirroring of axes, but it actually involves abstraction of the axes. Let's say that you have an upriver-downriver system, and you conventionalize it to the point that you would continue to call the same direction 'downriver,' even if at the point closest to the river from where you are, the river isn't actually flowing in that direction, or in the opposite direction, but in some other direction, right?

So if you conventionalized the system and abstract it such that the same bearings in absolute space will receive the same label regardless where in space you label them, regardless where you invoke them or refer to them—that's what Levinson calls an 'absolute' system, whereas systems are based concretely on landmark features that include what I call 'geomorphic' systems, a term that I borrow from Jackendoff, and 'landmark-based' systems. The latter are constituted by defining one axis of the coordinate system as a vector pointing toward or away from the anchor. If I say the presenter is 'toward the door from the bottle,' basically what I do is I compute a frame of reference by defining a vector that points from the bottle toward the door, letting that be an axis of the frame of reference, and then situating the presenter, the figure, in that frame of reference.

Levinson treats landmark-based and geomorphic systems as huge landscape-scale intrinsic frames of reference. I don't need to discuss the relative merits and disadvantages of this classification; what matters for our purposes is to understand that it is the classification that he proposed.

Now obviously the question is why is this interesting? The reason that it's interesting is that there is a surprising and in fact bewildering amount of variation across populations, across cultures, and across languages in terms of the frames of reference that people use to solve particular tasks, tasks of linguistic communication as well as tasks of spatial reasoning.

So the question is now how do we study this variation, these preferences. There are a number of tools which, in a more abstract way, I already talked about on Monday and Tuesday, but let's briefly review them. Ultimately everything should start with and end with linguistic data that actually are based on the observation of speech events as they occur in the community without the influence of the observer.

So that would mean spontaneous conversation, but since that's often too ambitious a goal, also staged recorded narratives, such as in particular local history narratives. Think about a narrative of how the village where you do

field work was founded, how, you know, people first started settling in this place, and so on and so forth. Of course, you know all of that would be situated in real space, and you'll get a lot of great, very productive material.

Very briefly, for Yucatec speakers, there are a lot of cultural practices that require the use of cardinal directions, absolute frames of reference based on the direction on the horizon on which the sun rises and sets.

This, from a sociolinguistic and culture perspective, this is very crucial information because it explains the striking gender bias in the use of frame of reference in Yucatec speakers. So it's almost exclusively male speakers who use the cardinal direction terms, because all these cultural practices that require them, such as house-building, or the horticultural gardens that people make in the jungle. Those are squares, and the edges of these squares have to be aligned with the cardinal directions, because if you don't do that, you invite bad winds that will spoil your harvest. So in other words, there are a lot of cultural practices that require the understanding of the cardinal directions, and those are predominantly, if not exclusively, in the male domain.

So if you can observe these events—whether you video tape them or just observe them with your own eyes, whatever is culturally acceptable—that's going to give you very rich material. Then you can do elicitation. The most productive approach, at least that has proven so over the years that my colleagues and I have been working on this, are those referential communication tasks. This concept of referential communication task comes out of the intersection of conversation analysis and psycholinguistics, from chiefly Herbert Clark and collaborators, who have been working on the cognitive processes involved in conversation for thirty years.

These tasks often involve something like a matching game, where in one way or another two stimuli have to be matched. For example, one has a photograph of a weird toy structure and instructs the other to build another copy of this toy structure, and the other person, the 'matcher,' has the toys to do so, and of course what's interesting from our perspective is the linguistic instructions that the 'director' is going to use. In order to force the director and the matcher to be maximally explicit about their spatial references, we put this screen between them to prevent them from sharing visual attention. So they don't have—or only in a very limited sense—a shared field of vision, which, of course, that's not going to work this way if you are going to work on sign languages. But people have found ways of overcoming that problem.

I already mentioned that an important and valid objection to referential communication tasks is ecological validity. This kind of task is very artificial. People don't generally use language in a setting where they don't share visual

attention. Increasingly for us that is no longer true, we constantly communicate via texting, chats, email, and so on and so forth. For most of the languages on the planet, they are spoken in interactions that people are having while they are in some place together, in real space, if you will, are still the norm. Of course, the way to overcome this problem is precisely to check whatever results you get out of a task like this against less controlled data: taping, recordings of spontaneous events, and staged narratives.

Let's consider the referential communication task that my collaborators and I developed for the project on the study of spatial representations in indigenous languages of Mesoamerica. It's based on an older stimulus that was developed at the Max Planck Institute in the 1990s. That older stimulus is the Men and Tree pictures, which are sets of pictures just like those that I'm going to show you in a moment, except those pictures feature a little toy man and a little toy tree, and what is variable across these pictures is the orientation and the location of the toy man vis-a-vis the toy tree.

So what we did is we got rid of the toy man and toy tree and replaced them with a ball and a chair, and these are photographs of a ball and a chair in real space. Why? Because one problem with using a toy man and a toy tree is that you get a bit of problem. You are forcing the speaker to make a bit of a tough choice when it comes to picking a figure and a ground, because the toy man is not a prototypical ground since he is animate. He is a toy, but he is supposed to represent an animate being.

Of course this is another problem that we tried to overcome, the fact that the Men and Tree stimuli used photographs of toys, meaning a representation of representations of stuff. We tried to at least cut one level of representation, one semiotic layer out of that.

We are still using photographs, which is still problematic, if speakers are used to using geocentric frames of reference or environmental frames of reference. So, let's say you want to say 'The ball is north of the chair' or 'The ball is east of the chair,' or something like that. That means you have to project the axes of the coordinate system from your real space into the space of the photograph, which is a little bit weird, partly because real space is three-dimensional and the photograph is two-dimensional. That's not the only problem.

So it's been suggested that use of photographic stimuli suppresses or depresses the use of geocentric frame of reference. I have some preliminary evidence to back that up, so we've developed an cognitive elicitation technique that uses three-dimensional objects rather than photographs, and I did get a higher percentage of usage of geocentric frames among the Yucatec speakers with that one than I got with this one, with the photographs.

Anyhow, back to the problem of figure and ground assignment. So the toy man is not particularly good as the ground, because he is a representation of an animate being. And the toy tree is not particularly good as the ground because it's not featured. It doesn't have an intrinsic front-back axis, therefore it doesn't project front and back, and left and right regions either, in intrinsic terms. It does in relative terms, but not in intrinsic terms. So what that means is, at least by hypothesis, the Men and Tree stimulus, in addition to suppressing geocentric frames of reference, will also suppress or depress intrinsic reference.

So what do the participants do? You have those two participants sitting side by side with a screen between them. Each of them has a set of 12 pictures in front of them. They are both looking at the same set of pictures, meaning two copies of the same set of pictures, except the pictures are in a different order on the two sides of the table. One participant, we call them the 'director,' is going to pick up a picture, describe it to the other, who we call the 'matcher,' and the matcher's task is to find the counterpart of that particular picture in their set. If they need additional information beside the description that the director starts out with, then they can ask questions.

The original set, the original Ball and Chair task, involves four different sets of 12 pictures each, and in the meantime we've added three more sets for a new line of research which addresses a research question which we didn't originally have on the radar screen.

Using tasks like these, we have identified the level of variation around the world. This is based on data that was collected by members of Levinson's group since the early 1990s, but also by the anthropologists Pierre Dasen and Jürg Wassmann and collaborators. A lot of it comes from my own project on spatial languages and cognition in Mesoamerica, which I will be talking about a lot today. Which is why you get a sort of cut-out of the Mesoamerican area here. So we have a lot of information on this area, because this is what my collaborators and I have been working on since 2008.

So what you see then here is that the blue dots are the languages the speakers of which prefer the relative frame of reference in small-scale space. Of course in terms of numbers of speakers, on a global scale, that's a huge portion of the pie, because it includes English, and probably most western European languages. It also includes Japanese, and probably a lot of other languages that are spoken in industrialized and highly literate languages, although this is in fact, you know, I mean, you notice that Mandarin is not on the map.

A year ago we started a second phase of this project, which was originally focused on the languages of Mesoamerica. The second phase now includes

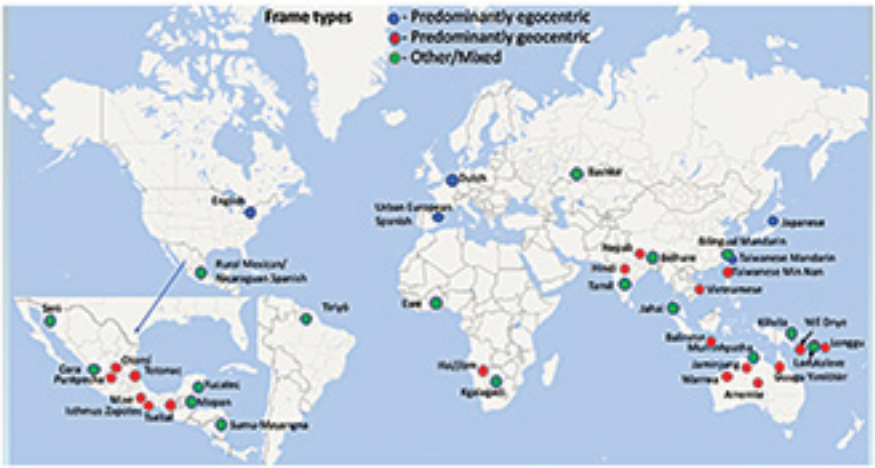

FIGURE 7.1 Reference frame use in small-scale horizontal space across languages
COPYRIGHT YEN-TING LIN. REPRINTED WITH PERMISSION

a bunch of languages from Asia and Africa and Oceania, and that includes Mandarin and Taiwanese. We have some preliminary evidence suggesting that intrinsic object-centered frames of reference play a much larger role in Mandarin and Vietnamese than they do in English or in Japanese, at least in Japanese that's spoken on Honshū.

So in terms of numbers of languages, it seems that this relative bias is actually a minority opinion. It's a fairly rare strategy, a fairly rare approach to dealing with space. And much more widespread are approaches that mix intrinsic frames of reference and geocentric frames of reference. The red circles are geocentric, and the green ones are the "anything goes" people. So that includes, for examples, speakers of the Kwa languages Ewe, spoken in Ghana and Togo, and that also includes Yucatec Maya speakers. So these people happily use all kinds of frames of reference. What that means is you get a lot of very complex utterances that combine multiple frames of reference, where people would say not just … let's say 'The bottle is in front of the chair,' but they would say 'The bottle is in front of the chair, to the north of it, behind it from my point of view, toward the window.' This is what Yucatec speakers do.

So this is very interesting because it raises some very interesting questions about spatial cognition. This leads us to the cognitive consequences. Based on the original research that John Haviland did on geocentric linguistic and gestural representations among the Guugu Yimithirr of the Cape York Peninsula in Hopevale in Australia, Levinson and colleagues developed the idea, the

prediction, that frames of reference that are preferred for communication should also determine the preferences for recall memory.

The reason is that you cannot translate information from one frame of reference into another. So let's say you memorize the position of the cat with respect to the car in relative terms. This huge eye is supposed to represent the perspective of the observer. So let's say you memorize the position of the cat as having been to the left of the car. So later on you went to talk to your neighbors about what you observed, but it so happens that the linguistic conventions of your community require you to talk about this scene in absolute terms. If you didn't memorize the location of the cat with respect to the car in cardinal terms, you are not going to be able to retrieve that information. It cannot be derived from knowledge of the location in the relative frame of reference alone, and vice versa.

So of course, one way to deal with this problem is to memorize the information in multiple frames of reference at once, and I'm sure that that's actually happening to some extent. It's an interesting problem for psychologists to study. I'm not quite sure to what extent that has happened. There is a lot of important research that was done in the early 1990s by Laura Carlson and collaborators on how people align different frames of reference, but I'm not aware of her having looked at recall memory.

So generally speaking it's a good assumption that if we can get away with mentally encoding a configuration in only one frame of reference rather than in multiple frames at once, we'll go for the simple thing, because it is less effortful. So that should mean that speakers of languages that prefer relative or egocentric frames in small-scale space will also memorize such configurations in relative or egocentric frames, whereas speakers of languages that prefer geocentric frames in small-scale space will use geocentric frames to memorize those configurations. And that is exactly what has been found.

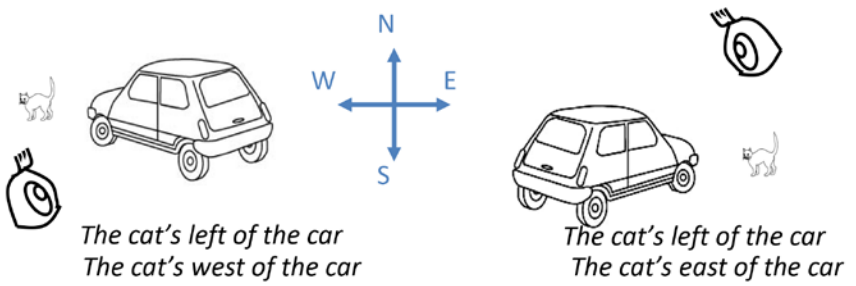

FIGURE 7.2 Limits of recodability across FoRs

COPYRIGHT STEPHEN C. LEVINSON. REPRINTED WITH PERMISSION

Here is the basic design of the most famous experiment that has been used to test this prediction, and all other experiments that have been used to test the same prediction. There is a large variety by now. But they all can be understood as involving a design that's a variation on this one.

You have an array of toy animals, and you put these in front of a participant, and you ask the participant to memorize this array, and their task is going to be to rebuild the array, so it is a memory experiment. You don't tell the participant what is going to be tested, is their memory for spatial information or just their memory for cows, and pigs, and sheep. It's a memory task.

When the speaker, the participant, tells you OK, they've memorized the array, you take it away, you wait for 30 seconds, then you walk the participant over to another table, which involves 180° rotation of the participant. You give them a set of four toy animals, including the original three, and you ask the participant to 'Make it again!' to rebuild the array as it occurred on the other table, on the first table, the stimulus table. But all you tell them is, 'OK, make it again!'

Now suppose you're me, suppose you are somebody who is used to thinking of small-scale space in terms of relative frames. So I'm going to memorize this array egocentrically. I'm going to memorize it in terms of, 'OK, they are all looking to the left, and they are forming a row, and the pig is in first position of that row, and let's say the sheep looks at the pig's butt, and the cow looks at the sheep's butt, which means the cow corresponds to my right shoulder, or is on my right of the array, and the pig is on my left of the array'. So I turn around 180°. 'OK, they are forming an array, they are looking left, so they are looking this way pointing, the sheep is in first, and it's on my left, the cow is in last, and it is on my right.' This is what Levinson calls the relative solution, although it is important to understand that strictly speaking it's not specifically a relative solution, but more generally an egocentric solution, because any egocentric encoding, even an egocentric intrinsic encoding, a 'direct' encoding in Eve Danziger's terms, would give you this solution.

On the other hand, if a speaker of Guugu Yimithirr or Hai//om, a Khoisan language spoken in the Kalahari of Namibia, or let's say, a speaker of Tseltal Maya in Mahosic in Chiapas is going to memorize the array, they will memorize the array as, let's say, facing south, and the sheep being the first, meaning the southernmost in the array, and the cow being the last, meaning the northernmost in the array. So 180° rotation, arrays facing south, sheep is the southernmost, cow is the northernmost, voila, there's your solution.

So the solution the speakers of these two languages will produce from memory are different. The speaker who uses egocentric frames of whatever kind will come up with one solution, whereas the participant who uses geocentric

frames of whatever kind, whether they are absolute or landmark-based or geomorphic, will come with a different one.

These are the responses from just two populations, namely, Tseltal speakers from Tenejapa, Chiapas, and Dutch speakers. Basically what you have on the *x*-axis is the percentage of absolute responses. Everybody did five trials, so 100% means five absolute responses. 60% means three absolute responses. 40% means two absolute responses. The *y*-axis gives you the number of participants that produced that number of absolute responses. So what you see is that nearly 90% of the Dutch participants produce not a single geocentric trial, whereas 60% of Mayan participants produce exclusively geocentric trials.

Let's consider participants from a much larger sample. So on the side of the linguistically relative populations, you have speakers of English, Dutch, Japanese, and urban Tamil, altogether 85 participants. On the side of the linguistically geocentric, populations, you have speakers of Arrernte, an aboriginal language of central Australia; Hai//om a Khoisan language from Namibia; Tzeltal Maya; Longgu, which is an Oceanic language; Belhare, a Tibeto-Burman language; and rural Tamil, altogether 99 participants. You can see a very similar distribution to what we have looked at before.

This is not the only cognitive consequence that has been hypothesized and in fact confirmed. A very important second cognitive correlate of the habituation to geocentric frames is that—especially if you are using large-scale geomorphic or abstract absolute systems, you really need to keep track of your bearings at all times. Because being a dedicated geocentric speaker means if you want to situate the bottle with respect to the chair, you have to do it in a geocentric frame of reference, so you have to know your bearings in geocentric space in order to be able to do that.

In other words, if your conventional way of talking about the spatial relation between the bottle and the chair is in terms of cardinal directions, north, south, east, west, as it is in many Australian languages, then that means you had better know, wherever you are, where north, east and south and west are. Unlike me, I never know this, I have to look for the sun, and if I can't see the sun, I am screwed.

So the prediction is that speakers of geocentric languages are much better 'dead-reckoners.' Dead-reckoners are people who keep track of how the orientation of their body changes in absolute or geocentric space as they move about. That prediction has been confirmed impressively in a type of task where you would take members of the population, you would take them out of their local community and into some neighboring town, put them in a building, a windowless building, spin them around a few times, and then ask them to point toward their hometown.

If you do that with Guugu Yimithirr speakers, it is incredibly accurate. Similarly here for these speakers of that Khoisan language of Namibia that I can't pronounce, these folks, actually, according to Levinson, narrowly beat homing pigeons.

In other words, the accuracy of their sense of absolute direction is slightly better than that of homing pigeons; which is stunning, because pigeons have an organ to do this. They have hardware built into their brains that allows them to keep track of their orientation with respect to the earth's magnetic field. Speakers of Guugu Yimithirr and that Khoisan language do it all with software, cognitively, computationally. In this case, software beats hardware.

These are responses from Tseltal speakers from Mahosic and you see that they all agree very much in where they are pointing, very much like the Guugu Yimithirr and the Khoisan speakers. They are all off by the same factor, and that's due to a difference in the orientation of the building, possibly also due to the effect of the local slope, since the Tseltal speakers use an uphill-downhill system. So the test was conducted in San Cristobal, I believe, or else in the capital of the county of Tenejapa, and something about the local environment messed with these participants' frames of reference.

I want to talk about two more recent studies, and show you a little clip. Daniel Haun and colleagues have produced some preliminary evidence suggesting that non-human primates have a preference for geocentric rather than egocentric cognition. Which is very important because in the developmental

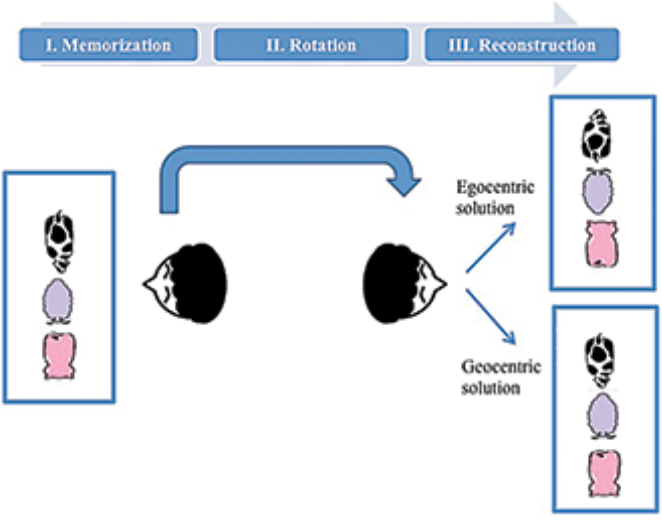

FIGURE 7.3 Design of array reconstruction tasks
COPYRIGHT YEN-TING LIN. REPRODUCED WITH PERMISSION

literature, it was for a very long time assumed that children the world over all have to master egocentric cognition, because that is just the way it is. That was assumed to be just innately built into the human mind, and this data, looking at the back-story in hominid evolution, produces more evidence that that earlier assumption was really not true.

I also mentioned another more recent study that showed that whatever frames of reference people are accustomed to—even if we are talking about young children, so they haven't been accustomed to this for very long—it's hard to break the habit, and it's hard to use a kind of frame of reference that you are not accustomed to, that you are not habituated to.

Now I want to show you a little clip. This is from a study that Daniel Haun and Christian Rapold did, again, in Namibia, and what you are going to see … I like to call this "Dances with the Anthropologist"—a play on the title of the Kevin Costner movie *Dances with Wolves*. So this is Christian Rapold, and he is going to teach this kid some dance moves, and then he's going to turn him around, and ask him to reproduce these dance moves, and watch what happens.

So the kid learned the dance moves in geocentric terms. That was the kid's interpretation of how the instructions were intended, which, you know, it's quite amazing—this kid uses a geocentric frame to memorize movement of his own body.

I have presented some evidence to the effect that there is an alignment between preferences for particular types of reference frames in discourse and preferences in non-linguistic cognition, particularly in recall memory, but it's also been shown to obtain for spatial inferences. The question is now how do we interpret this alignment? Two interpretations have been suggested. The one that was proposed in the 1990s by Levinson and Pederson and colleagues is a Whorfian or neo-Whorfian interpretation. In other words, they are arguing that language is the driving force. It's not the exclusive factor but it's the strongest factor. On the other hand, Peggy Li and Lila Gleitman published a paper in Cognition in 2002 in which they argued that language may not be a factor at all.

From their point of view—and this is where the main difference lies, the main disagreement—Li and Gleitman assume that everybody, regardless of which population they are socialized into and which languages they speak, has the same innate knowledge of all types of reference frames. Now think about that for a moment, because what that means is if you assume that a particular type of upriver-downriver system is conventional in a particular community, and the way it's used in that community constitute a particular type of reference frame, that will mean everybody else in the world will have to have innate knowledge of that type of reference frame.

That seems fairly absurd, but of course, these authors would then claim that, well actually this conventional type—that is based on the local environment in a particular community and on conventions among the members of this community that are based on the local environment—is not really a separate type of reference frame. This is something that one can easily pick up just by going there and looking at the local environment.

So in other words, the preferences, the variation in preferences that have been observed across populations is supposed to be just a matter of the adaptation to factors of the local environment, such as, is there some sort of salient object or entity in the environment that can serve for projecting a particular reference frame type, and factors such as literacy and education and infrastructure and population density. In other words, those environmental and demographic factors are supposed to drive preferences for the use of spatial frames of reference in both discourse and non-linguistic cognition. Language plays no role in the cultural transmission of practices of spatial reference, because actually there is no cultural transmission of practices of spatial reference for these authors, because from their point of view, everything happens in the mind of the individual member of the community in response to those environmental and demographic pressures. So they are assuming this version of the big picture where the important parts of cognition, namely the knowledge and the ability to use different types of reference frames, are innate and the role of language in internal cognition is negligible.

On the other hand, Levinson and colleagues assume this version of the big picture where the reference frame types are not innate, and of course, the research of Haun and colleagues on non-human primates strongly backs this position. They are not innate. They are learned. The computational, cognitive ingredients that we need in order to form reference frames, those might be innate. Meaning the ability to identify the axes of objects and geometric operations, such as translation and reflection—those things might be innate. Or maybe they are not, but it's possible. But the reference frame types themselves are conventional, and they are learned through cultural transmission.

That means you then are faced with a paradox that becomes clear once we take another look at this map from earlier. So you see that across human populations, there's a bewildering array of strategies, there's a great deal of variation. However, within many of these communities, there is much less variation. For example, speakers of Isthmus Zapotec agree to a very large extent on the use of cardinal directions in small-scale space. So this is what upwards of 80% of all the adult respondents do. You get the same figure for the preference for relative frames of reference among Dutch speakers. So that means there is much less variation inside many populations than there is across populations.

The question is how is that possible? What it means is there must be some way for the members of each community to converge on the solution, the set of preferences, the usage profile that is conventional for that community. And how do they do this? This is again the problem that I talked about on Monday. We are generally not very good at mind reading, telepathy. One could claim we suck at it. Most of us do, anyways. That's why we have language and gesture to communicate amongst one and another, and we use language and gesture to communicate and thereby transmit not just our verbal messages, but also the styles of thinking that go into computing these verbal messages.

In other words, children observe not just the utterances and figure out their meanings, but they also are able to infer the underlying practices of spatial reference, including conventions for reference frame use, from language. Not solely from language. They are going to get a lot from gesture. They are going to get a lot from observing other practices, such as, you know, those cultural practices that I mentioned that tuned Yucatec speakers into the use of cardinal frames of reference.

But language is going to play an important role in this cultural transmission. Simply because it's the most powerful system or modality that we have at our disposal for transmitting information among one and another. So in other words, from the perspective of the neo-Whorfians—and that's my perspective too—the principal reason why language is bound to be a causal ingredient in the preferences for reference frame use is because it serves an important role in the cultural transmission of preferences for using particular types of frame of reference, or generally in styles of cognition, cognitive practices.

Let's consider a paper Asifa Majid and colleagues published in 2004 where they did an initial test of Li and Gleitman's claim that there is a correlation between particular usage profiles and cultural traits, and they found that that claim was difficult to find evidence for. There is an important exception, which is that the preference for relative frames of reference has been attested so far mostly in highly literate and educated populations.

So literacy and education are factors alternative to language that still stand a fighting chance of being driving factors that explain differences across populations. This is very much something we're currently addressing in this project on spatial frames of reference, or spatial language and cognition in Mesoamerica and beyond.

Very briefly, let's consider some empirical tests that have been conducted by Li and Gleitman. So basically, remember that what they are trying to show was that whatever preferences exist across populations are superficial, and it can easily be changed because fundamentally cognitively we are all the same. Everybody has innate knowledge of all types of reference frames.

So in 2002 they did experiments with American college students and tried to show that American college students can be induced by relatively pedestrian environmental manipulations to use geocentric rather than egocentric frames of reference. One thing they did was test those college students indoors versus outdoors, and the participants turned out to produce more geocentric responses when tested outdoors.

That actually did not replicate when Levinson and colleagues tried to replicate it in the Netherlands. So you know it's unclear why Li and Gleitman got this result and Levinson and colleagues did not. But Li and Gleitman was alleging that mainly a lot of the research that have been done with indigenous populations have been carried out outdoors, and that's why people got more geocentric responses. But that's actually not true. Almost none of the studies that contributed to that map that I showed you were carried out outdoors.

Another thing they tried to do is, they replicated this 'Animals-in-a-Row' task, meaning what Li and Gleitman did. This time they put a toy pond on the stimulus table and on the demonstration table, and of course, whichever way or wherever they put the toy pond, the participants would array the animals such that they would be facing the toy pond. And Li and Gleitman claimed that that means they were using a geocentric frame of reference.

But of course, in reality, what the participants really were doing was memorizing this entire array globally with the orientation of the animals toward the toy pond being an intrinsic feature of that array. To show this, Levinson and colleagues replicated Li and Gleitman's duck pond condition using 90° rather than 180° rotation, and of course, that led to people going not with the egocentric response. So imagine we have a toy pond on the stimulus table. This is the toy pond on the record table, and of course the responses that people went with is this one rather than this one, which would have been the absolute response, or the geocentric response. Showing that the frame of reference that the participants used to encode this array was not in fact a geocentric one but an intrinsic one. Typologically, the intrinsic frame of reference is supposedly available to speakers of all languages, with the possible exception of Guugu Yimithirr. Therefore it's not surprising that English and Dutch speakers use intrinsic frames of reference to solve a problem like this.

More recently, Li, Abarbanell, Gleitman and Papafragou published a paper in which they tried to, this time, show that with similar small manipulations, Mayan people can be induced to use relative frames of reference. And there the issue is basically that the tasks they used do not in fact rely on true relative frames of reference, but merely on intrinsic direct egocentric frames of reference.

Let's say one of the tasks they did is they used playing cards with geometric patterns on them, which had colored circles on them. So the participant would be shown a particular card at the stimulus table, and then this card would be put in a box. The participant would be holding the box, then they would move over to the recall table and without opening the box would have to identify the counterpart of the card on the recall table.

There are two conditions, one under which the participant carries the box with the original cards such that they rotate it along with their body, preserving the orientation in egocentric terms. Under the other condition, they have to preserve the orientation of the box with respect to the room in geocentric terms. The results show that the speakers had about equally frequently success under the egocentric and under the geocentric condition, which Li and Gleitman claim shows that these participants are just as adept at using the relative frame of reference as they are using geocentric or absolute frame of reference.

The problem with this is once again you can solve the task simply by memorizing the original card with respect to your own body, not just as anchor, but also as reference entity. So you are going to memorize it along the lines of, 'OK, the red dot is near my left hand, and the green dot is near my right hand.' You don't actually need transposition in order to do this. You don't need a relative frame of reference. You can do it with an intrinsic egocentric frame of reference, and those are again typologically not predicted to be restricted across populations.

This is an ongoing debate. I want to quickly introduce the project on spatial language and cognition in Mesoamerica, which I've already mentioned several times, which is precisely dedicated to advancing this debate, and I'll present the analysis that we've been trying to do in order to make some progress on Li and Gleitman's hypothesis. Meaning, you know, the hypothesis that it's not language that is a causal factor in population-specific preferences in spatial cognition. Rather, these preferences are simply the result of cultural adaptation.

So this is what we've been trying to get a grip on. And I'm going to show you some preliminary data. Because that's going to involve quantitative analysis, so that's why I'll put it in the lecture on quantitative methods in semantic typology. Right now I just want to introduce the project.

The original project involves fifteen different field workers working on thirteen Mesoamerican languages, four Mayan languages, three Mixe-Zoquean languages, two Otomanguean languages, and Tarascan, or Purepecha, which is a linguistic isolate; Huehuetla Tepehua, which is a member of the very small Totonac language family; and then two Uto-Aztecan languages, one of which, Cora, you could consider as being situated on the very edge of Mesoamerica,

so in some way it's categorized as Mesoamerican language and in some way it is not.

Then we have two indigenous 'controls' spoken just outside the Mesoamerican area: Seri spoken a few hundred kilometers north of the Mesoamerican area in the state of Sonora, and Sumu-Mayangna spoken in Nicaragua just south of the Mesoamerican area. And in addition, this says Mexican Spanish, but by now we have been looking at three different Spanish varieties: Mexican, Nicaraguan, and European Spanish from Barcelona.

The project is dedicated to two interrelated domains. One is spatial frames of reference. The other is meronyms, meaning linguistic terms for object parts, including body parts, which are known to be highly productive in Mesoamerican languages. And one interest is in the underlying cognitive processes involved in assigning meronyms, but another interest is in finding out to what extent the use of these productive meronyms as a resource in spatial reference influences the use of spatial frames of reference.

So we have a hypothesis according to which speakers that rely on geometric meronyms in order to individuate what I was referring to in the beginning of the lecture as 'place functions,' using Jackendoff's term, that speakers of such languages would not prefer relative frames of reference.

And a year ago, the second stage of this project started, although the old one continues, so currently both phases are running side by side, which gets a little complicated once in a while. So we've been expanding, looking at a bunch of languages from outside the Mesoamerican area, including Mandarin and Taiwanese. Alright, I'm going to skip this and move to brief conclusions, and then we'll have a little bit of time for discussion.

So the Linguistic Relativity Hypothesis is the hypothesis that linguistic categories determine categorization … There is something missing here, that linguistic categories determine nonlinguistic categorization, it should say. That's the strong formulation. Or that linguistic categories may influence non-linguistic categorization, that's the weak formulation. And the distinction between the strong and the weak version of the Linguistic Relativity Hypothesis goes back to Roger Brown, who introduced this in his obituary for Eric Lenneberg in 1976.

Spatial frames of reference are conceptual coordinate systems used to identify places, orientations, and directions in discourse and internal cognition. We talked about the debate on linguistic versus nonlinguistic factors that influence the use of frames of reference. We've seen that different populations prefer different reference frames for the same task and domain. These population-specific preferences align for discourse and internal cognition, such as recall memory and strategies of spatial inference.

Levinson and Pederson and colleagues have been arguing that language is the, or a, driving force, a causal factor, an important causal factor, let's put it that way, whereas Li et al. (2011) argued that variation across populations is in fact the result of adaptations to the local environment.

Of course these adaptations are real and nobody denies that. Clearly the reasons that Tseltal speakers use up-hill-downhill terms is because they live in the mountains. No doubt about it. Where the two sides disagree is in whether such adaptations happen in the mind of the individual speaker at an ontogenetic level, or whether they happen culturally at a phylogenetic level. So the neo-Whorfians assume that these adaptations themselves are conventional in the culture, and rather than to come up with them individually, one by one, the members of the community learn these adaptations from one another culturally.

LECTURE 8

Doing the Math: Quantitative Methods in Semantic Typology

This lecture will be about the quantitative side of semantic typology, although in a sense that's misleading, because as I'm going to point out in a second, typology is inherently a quantitative enterprise. What we are going to look at in this lecture is more specifically the use of statistical methods in typology, in semantic typology in particular.

I have got to start with a disclaimer simply because I don't actually know very much about statistics, and most of what I do know about statistics I've learned in the last two years. I'm going to present mostly studies that other people have done, have published in recent years, and I'll tell you what I think about those, and in the end I'll show you some work-in-progress that my colleagues and I have been pursuing in this project on spatial language and cognition in Mesoamerica that I introduced this morning. So in a sense, that final part here, that fourth study is the follow up and the conclusion to the cliffhanger that I left you with this morning.

And then finally I'm going to ask the question: are we actually ready to go quantitative? And what I mean by that is, semantic typology, and typology at large, is now embracing statistical methods, but is doing so while analyzing data that have been collected according to the humanities standards of the past, which leads to funny mismatches. This is something that we need to be aware of, and need to try to remedy as we go along. So we are not yet in some sense doing serious quantitative research in my view, and we must understand that.

Typology—not just semantic typology, but any form of linguistic typology—focuses on the distribution of linguistic properties across the languages of the world and aims for generalizations of the type 'All languages have feature X,' 'Some languages have feature X,' or 'If a language has feature X, it also has

All original audio-recordings and other supplementary material, such as any hand-outs and powerpoint presentations for the lecture series, have been made available online and are referenced via unique DOI numbers on the website www.figshare.com. They may be accessed via this QR code and the following dynamic link: https://doi.org/10.6084/m9.figshare.11419137

© JÜRGEN BOHNEMEYER, 2021 | DOI:10.1163/9789004362628_009

feature Y.' This third type is by far the most common generalization that can actually be supported by typological data. The second type in a sense is not necessarily very interesting, and the first one is almost never borne out, because for any kind of strong universal of this type you can always find exceptions, as Evans and Levinson (2010) pointed out.

Why is research that is aiming for this kind of generalizations quantitative research? Because these are quantified statements. They involve logical quantifiers. They talk about all languages or some languages or they have a conditional generalization that is supposed to be valid for the entire set of extant natural human languages. Now there is a difference obviously between quantification in the logical sense and quantification where actual numbers are involved, but there's also a close and obvious relationship.

The reason typologists have found they need to play the numbers game rather than merely go for logical quantifiers is because going for the logical quantifiers often doesn't give you very much information about the distribution that can be found in the languages of the world, because these distributions tend to be fairly complex. Moreover, there is the very substantial problem that we can not actually survey the entire set of languages spoken on the planet today at all, not even 50% of them or 20% of them. We have to make do with samples of languages that are much smaller. For that reason also, it is increasingly recognized as important that we quantify our generalizations.

So the use of statistics in typology goes back to the 1980s and was pioneered by Joan Bybee and Matthew Dryer in syntactic typology. These authors used primarily methods of inferential statistics, meaning they tried to test whether an observed distribution in a given large sample of languages—and when I say large, I mean hundreds of languages, databases comprising data from hundreds of languages—where a given distribution would be significantly versus not significantly different from chance.

But in recent years, there has been increasing use of methods of descriptive multivariate statistics and also inferential multivariate statistics, which are approaches that allow you to break down an observed distributional data set and find the factors that govern this distribution. I know this sounds a little bit mysterious—I'll try to explain as best as I can as we go along.

This has been going on in syntactic typology as well as in semantic typology and phonological typology, and it's become a trend *not* because these are new methods. These methods of multivariate, primarily descriptive statistics—descriptive in the sense of techniques that simply allow you to represent the distribution of the data in a way that is a little more insightful than just looking at a spreadsheet, if you will, as opposed to inferential statistics, which allows you to determine how likely a given distribution is different from chance. The

reason there has been a boom in such methods in recent years is not because these techniques are new. The techniques have been around for many decades. Rather, it's that nowadays even linguists can afford to actually apply these kinds of methods, because the software that you need to do that and the knowledge that you need to use the software and the computational power that you need to run the software has become so cheap that anybody can do that.

And when I say anybody, I mean there are students in my lab who can do this. Apparently I'm already too old for this kind of thing. I'm 47 and when I was a student I never took a statistics course, because it was not obvious to me back then I would ever need that kind of thing for doing linguistics. Nowadays I could kick myself.

In the last two years there has been a series of publications in general science journals, such as in *Science*, which had an article by Atkinson on apparent founder effects in phoneme systems; the article in *Nature* by Michael Dunn and colleagues using phylogenetic statistics in application to word order patterns, trying to show that these are by and large inherited from ancestral languages and not in a demonstrable way shaped by universal pressures such as cognitive factors, processing factors, according to the authors. I mentioned a brand-new article by Charles Kemp and Terry Regier on the communicative versus cognitive efficiency of kinship term categories across the languages of the world. That would be another example. That just came out two weeks ago in *Science*.

So there's a real trend here. And you see that this trend actually possibly allows linguists almost for the first time to enter a general science fray and, you know, make news, make headlines, that scientists outside linguistics and even outside the cognitive sciences notice and pay attention to. Unfortunately, one sometimes wishes that a few more linguists would be involved in the review process before these kinds of things come out in *Science* or *Nature* or *Proceedings of the National Academy of Sciences*.

The first application of descriptive multivariate statistics, in this case multidimensional scaling, to semantic typology was an article by Stephen Levinson and Sergio Meira and the Language and Cognition Group, of which I was a member, that was published in *Language* in 2003.

So we'll start with Levinson and Meira, which was the first use of multidimensional scaling or any descriptive multivariate statistics in semantic typology in a narrow sense. It is actually the case that some anthropologists already used to use similar techniques of analysis a long time ago, decades ago, but not in explicitly typological research.

Levinson and Meira's study is based on the data that Melissa Bowerman and Eric Pederson got a large number of us to collect from speakers of languages

around the world using the Topological Relations Picture Series. And the aim of that paper was to use, to test, what Levinson and Meira called 'orthodox assumptions' about topological spatial relations.

Now, those of you who attended the morning lecture will remember that topological spatial relations are perspective-free spatial relations. They are relations that do not depend on a frame of reference, in other words. So if I say that 'The cap is on the bottle,' that is going to remain true even if I invert the bottle. It does not depend on whether I look at it from my point of view or your point of view. It does not depend on where the bottle is in absolute space. That's basically what Piaget and Inhelder, way back when, defined as 'topological' spatial relation. Their idea was that children learn these before they learn the projective relations, the one that depend on a frame of reference, because the topological relations would be conceptually simple.

There has been an old assumption in the literature based on English that there are three primitive topological relations: AT, ON, and IN, which more or less corresponded to the meanings of the English prepositions *at*, *on*, and *in*.

AT would be used primarily with one-dimensional grounds or with grounds whose spatial extension and shape does not matter to the selection of a spatial relator except for that one dimension. That's specifically the case when you're trying to express a relation of proximity or perhaps a relation of a zone of dominance, a region of space that's defined as a zone dominated by a center, something like that.

So then you have ON, which occurs with two-dimensional grounds and typically with the relation of surface contact. And you have IN which occurs with three-dimensional grounds or two-dimensional enclosures and which involves the relation of containment. And of course you can combine these with locative predicates or even with semantic locative functions, incorporated in adpositions, depending on the linguistic type that you're dealing with, but also with goal or allative functions, with source and ablative functions, and with route functions.

So then you get this set of prepositions of English, all of which are based on the supposedly primitive AT, ON, and IN relations. So you have *at the station*, *to the station*, *from the station*, and *via the station*, all of which would involve the topological function AT. And *on the table*, *onto the table*, *off the table*, and *across the table*, all of which would involve the primitive topological relator ON, and so on and so forth. And this concept goes back to a paper by Herb Clark, one of his earliest, published in 1973, and the main point there is to make predictions about the learning of these prepositions in child language.

A set of orthodox assumptions is that there are these primitive conceptual relators, AT, ON, and IN, and that these across languages will tend to more or

less homomorphically map onto the meanings of linguistic expressions, such as prepositions in English, postpositions in other languages, and yet other classes of expressions in yet other languages. So that is the hypothesis that Levinson and Meira wanted to test.

I don't need to introduce the BowPed pictures anymore, but to recap briefly, it's a set of 71 line drawings which was originally designed by Bowerman and Pederson to study the classification of specifically relations that involve prototypically or more peripherally either support configurations or containment configurations. So the idea was to explore a conceptual space that centrally includes relations of support and containment and see how this space is structured across languages.

Crucially we talked about etic grids and the role of etic grids in the design of research in semantic typology. Bowerman and Pederson did not have an absolutely explicit etic grid for this study. They started out just wanting to do exploration and then the thing basically took on a life of its own from what I understand. Which means now in retrospect we are sort of forced to take the whole set of pictures, the 71 pictures, as representing an etic grid that was apparently never intended as an etic grid in this way, which is a somewhat interesting situation and has implications for this study that I'll get back to in a second.

This is the procedure, which I did talk about on Tuesday: so basically every picture shows a figure and a ground. The figure is identified by an arrow, and you have to get the participants, the speakers of the target language, to try to describe the location of the figure with respect to the ground in whatever ways are natural in their native language.

Now Levinson and Meira got those of us who contributed to this study to code their data for them focusing in particular on information in what I call the 'ground phrase.' That's a technical term, which I use for a constituent of spatial descriptions that is a sister, or complement, of the head of the predicate and that in its turn may either dominate the descriptor of the referential ground or may in fact be identical to the descriptor of the referential ground.

So in English, the ground phrase is a prepositional phrase. In Finnish, the ground phrase is often a case-marked noun phrase. In Yucatec, as we are going to see, the ground phrase is a noun phrase, although it's a noun phrase typically headed by a meronym, a relational noun. And there are many other possibilities across the languages of the world.

Levinson and Meira got us to code the information expressed in the ground phrase of the utterances that we collected, and specifically to identify a topological relator, which Levinson and Meira call the 'spatial adposition,' where 'adposition' is a cover term for preposition and postposition. And they defined

this notion as follows: "A spatial adposition is any expression that heads an adverbial phrase of location in answers to where-questions." "This isn't necessarily an adverbial phrase. In fact typically it is not actually an adverbial phrase. This definition is not designed to exclude spatial nominals, since they so often gradually develop into 'true' adpositions that boundary problems would plague a comparative exercise of this sort" (Levinson and Meira 2003: 486).[1] In other words, they tried to do an analysis over spatial relators that include both adpositions and meronyms and whatever other type of expression may occur. This gives you an idea of the grammaticalization process that may lead from a meronym to an abstract spatial relator, in this case from the back of a person to the back of a car, to the region at the back of the car, to the region behind the car.

So the sample of languages that they looked at comprises nine languages, all of which are genealogically unrelated, and which are spoken across a wide variety of different areas. Basque and Dutch are obviously spoken in Europe; Lao is the only Asian language; Lavukaleve and Yélî Dnye are both spoken in Melanesia; Ewe in West Africa; Tiriyó in Suriname and Brazil, Trumai also in Brazil, in the Amazon; and Yucatec is a Mayan language obviously.

Now one thing you notice here is that the data in the case of Basque comes from 26 speakers, in the case of Dutch and Tiriyó from 10 speakers each, and on the other hand, in the case of Lavukaleve from just 1 speaker. So this introduces an important comparability problem, and the way Levinson and Meira try to deal with that is they average the responses across the participants in a way that we'll see in a moment.

Basically what they did is they tried to aggregate responses across speakers and thereby of course completely eliminated any representation of inter-speaker variation in the data. So that's a significant problem in my view.

So the next problem is that a language may not just use adpositions alone, but in fact combinations of adpositions and meronyms, relational nouns in the ground phrase. And this is problematic, because Levinson and Meira decided to, in this case, encode the adposition only and ignore the relational noun.

So, let's see, here is a Yucatec example. This is a description—so the BowPed picture you see here. This is a dog sitting next to his doghouse, and the description [is] [(10)]:

1 Levinson, S. C., S. Meira, and The Language and Cognition Group. 2003. 'Natural concepts' in the spatial topological domain—adpositional meanings in crosslinguistic perspective: an exercise in semantic typology. *Language* 79(3): 485–516.

(10) *Te'l kul-ukbal u=pèek'-il tu=pàach le=nah=o'*
there sit-DIS(B3) A3=dog-REL PREP:A3=back DET=house=D2
'There the dog is sitting outside the house.'

So you have a preposition here, which is a generic preposition *ti'*, amalgamated with the clitic 3rd-singular pronoun *u* which marks the possessor of the meronym *pàach* 'back.' It's one of only two prepositions that occur in this language. And *ti'* really is semantically pretty much empty. You can translate it as 'with respect to.' It doesn't have a spatial meaning. Then you have a relational noun *pàach*, which means 'back.' However, when you use that as a meronym in spatial descriptions in Yucatec and other Mayan languages, it refers not just to the back, but also to the outside of the object. So what (10) basically means, the whole thing, is that the dog is sitting outside the doghouse.

Now Levinson and Meira's strategy for how they analyze the data was in case there's a combination of multiple relators, just go with the adposition. So the only thing the data got coded for in this case would be the adpositions, which means this description would be assigned to the extension of the preposition *ti*, and the fact that the relator *pàach*, which in this case gives you the outside region, occurs in the description was ignored. So that obviously is a problem, and we are going to see a very similar problem come up in the second study that I'll talk about in a moment, the study on verbs of 'cutting' and 'breaking.'

But you know, I am not trying to simply criticize the authors here. I want to point out that this is a problem that you are going to face when you are trying to do a quantitative analysis of this type of data. You are going to make decisions ultimately in how to categorize the data—that's what coding basically means—how to categorize the data on the basis of the expressions that occur in the response, in each response, and you have to sort the responses into bins, into categories, and that may often involve highly nontrivial decisions. So this is, you know, a general problem that comes up again and again when you are trying to quantify over this type of data.

These languages differ dramatically in terms of the number of relators they have, in particular the number of adpositions. So, Tiriyó, a Caribbean language, has more than 100. On the other hand, Yucatec has only one or two. And all the other languages are somewhere in between. But of course the fact that Yucatec has only one or two adpositions is partly due to the fact that a lot of the information that otherwise would be expressed in adpositions gets expressed in those relational nouns, those meronyms.

So this is just an attempt by the authors to show that there is no possible arrangement of the 71 pictures such that you could group the responses, the categorizations of the nine languages against this particular arrangement and get

only contiguous Venn diagrams, contiguous and circuit areas, to represent the extensions of the categories of these nine languages. So there is no single arrangement, no single two-dimensional arrangement of the scenes that would allow you to do that, which is just a way of saying the way this space is broken down across these 9 languages is pretty different.

But in order to find out just how different this is, we need some more sophisticated analysis, and the answer to that is multivariate statistics. So the technique that Levinson and Meira used is Multi-Dimensional Scaling. Multi-Dimensional Scaling and a host of other similar techniques, such as Cluster Analysis, Factor Analysis, Correspondence Analysis—all of these, as well as other techniques, phylogenetic techniques which I will introduce later today are based on the analysis of similarity matrices or distributional matrices.

So what's a similarity matrix? A similarity matrix tells you about a set of observations, abstractly speaking, how similar they are to one another. That is, it's a grid that compares every observation to every other observation and tells you how similar are these to one another in the outcomes, in the way they were treated by the participants. Now when I say 'observation,' that can mean different things. You can group the data by the stimulus items involved in the research. You can group the data by the languages. You can group the data by the participants. There are different ways of constructing a similarity matrix. We'll see examples of that as we go along.

In the case of Levinson and Meira, they compiled a similarity matrix over the stimulus items, meaning over the 71 pictures. So what that means is, imagine a two-dimensional grid where we compare each of the 71 pictures, so this is picture 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and so on until we get to 71. We compare every picture to every other picture, and we ask the question across the languages of our sample, how similarly are these pictures treated to one another? That's a similarity matrix.

Now, the way Levinson and Meira calculated their similarity matrix is, they first asked for each language which of the BowPed pictures are in the extension of the same adposition. I mean adposition in the sense that for the purposes of this analysis, the strategy was identify one topological relator and treat that as the adposition, whether it's actually an adposition or relational noun or whatever.

So, find the number of adpositions for the given language that are shared between the two pictures you want to compare, subtract that from the total number of adpositions that occur in the responses from this language, and divide the whole thing by the number of total adpositions of that language. And so that means, the more adpositions two pictures share, the more similar

they are to one another. And the fewer they share, the more dissimilar they are to one another.

And so in other words, if two pictures are in the extension of one single adposition, they have a given degree of similarity. But if they are in the extension of more than one preposition, if more than one preposition or postposition in the same language can be used to refer to both of these items, then the similarity increases. So that gives you a similarity matrix for one language. Then you simply add up the similarity matrices for the nine languages that constitute your sample, and the result is a composite similarity matrix. Now, you can apply a multi-dimensional scaling algorithm to this similarity matrix.

And what does a Multi-Dimensional Scaling (MDS) algorithm do? It basically creates a two-dimensional plot of the data in the similarity matrix.

So suppose you wind up with a measure of similarity for each of these pictures. So of course every picture is similar to itself, so that's going to be 1, and so on and so forth. And then let's say here you have 0.5. Here you have 0.47, 0.3. Maybe up here you get 0.8 or something like that, who cares. So then notice what you are getting here you can interpret as a set of distances. You know these tables that will show you distances between cities in a country or on a continent. So let's say you have Beijing, Shanghai, Hong Kong on the x-axis, and Beijing, Shanghai, Hong Kong on the y-axis. And obviously from Beijing to Beijing is zero kilometers. From Beijing to Shanghai it's maybe 938 kilometers. From Beijing to Hong Kong, it's, I have no idea, I'm going to say 875, and so on and so forth. You can compile a matrix of these distances.

Suppose now what you do is you want to actually draw a map of these different cities, but all you have to go by is the matrix of the distances. So you're looking for the simplest spatial model that fits these matrices of distances, and that's exactly what we are going to do. We are going to interpret these similarity indices as distances, and we are going to create a two-dimensional plot of all the pictures of the similarity space, showing the location of all the pictures in the stimulus kit, which is based on these distances. The distances are a representation of the similarity of the pictures in terms of the way they're classified across languages.

And this is what you get when you do that: a two-dimensional visual representation in which the similarity indices are interpreted as Euclidean distances between the pictures. Euclidean distances in the sense that if you have two points with coordinates, then the sum of the square of the differences on each axis gives you the sum of the square of the distance between the two points.

If you want a spatial model that loses no information, then given you want to represent distances among 71 points, you might want to need an up

to 70-dimensional space to get a fair representation, a good spatial model. So what you are doing when you are forcing this model into just two dimensions is you are producing a lot of stress, you are in a sense deforming the spatial model—and that means you are losing a lot of information.

But on the other hand, this should allow you to identify the two most important conceptual dimensions that structure the similarity space. Now that's not something that Levinson and Meira did. They did not try to interpret the dimensions of this two-dimensional plot, but what they did do is interpret the clusters. So you notice that the 71 pictures, in terms of how they fall in this two-dimensional representation of the similarity space, based on how they are semantically categorized across the nine languages, fall into five different clusters. Then there're a lot of pictures that don't really fall into any cluster, but are sort of asteroids out in outer space.

The analysis that Levinson and Meira present with respect to this plot is they're saying, look, if it was really the case that there are these three conceptual primitives, AT, ON, and IN, and those are recurrent across languages, may be universal, may be innate—then we should find three clusters in this data set that correspond to these three notional topological relators AT, ON, and IN. However, in reality what we are finding is there is an IN cluster. There is also a NEAR cluster, but that's not simply AT, because it also includes a number of UNDER scenes. But there are a few scenes where you have a figure under ground. There is also a large cluster of attachment scenes, which you could classify as either AT or ON in English, depending on what description you pick—and it is not a very sharp cluster, so this is sort of unwieldy. Then there is a cluster that includes superposition scenes which could be treated either in terms of ON or in terms of OVER.

Levinson and Meira conclude [that] the hypothesis of this kind of set of universal primitive topological relators AT, ON, and IN, failed to confirm. In reality, what we are finding is evidence of a much more complex structure of the notional space underlying topological relators across languages.

In addition, they try to come up with a sort of evolutionary scale along the lines of Berlin and Kay's evolutionary scale of basic color terms, along the lines of, if a language has a relator for ON—meaning a relator that is dedicated to surface contact or support—it will also have a relator that is dedicated to containment, and so on and so forth. And the problem I have here with that is the Spanish preposition *en*, which I mentioned before on Monday and Tuesday, which covers AT configurations, IN configurations, and ON configurations. So if there is such a thing as a generic AT category that is the simplest kind of preposition you have, then Spanish *en* should come pretty close to it. And yet,

en coexists with a very large number of semantically more specific adpositions, which does not conform to this idea of an evolutionary scale. So it's not obvious that adpositional systems really submit to this kind of analysis. But then, you know, obviously, if that's indeed not the case, then we have to continue to ask, why is it not the case? Why is it apparently the case that basic color terms submit to this kind of scale, but adposition systems do not?

Nevertheless, my main point is what I call the stimulus artifact problem. So I have two main objections against this study. One was the aggregation, the way they averaged the responses across participants, thereby wiping out the effect of inter-speaker variation—for example giving responses from just a single Lavukaleve speaker an equal weight in how they affect the final outcome of the analysis to responses that come from 20 odd Basque speakers.

Secondly, the stimulus artifact problem, meaning the fact that BowPed was never designed as a tool for the study of the entire semantic space of topological spatial relations. But that's the way Levinson and Meira treated it. In reality, BowPed was only designed to study this continuum—or at least Bowerman and Pederson assumed it would be a continuum—of relations between support and surface contact and inclusion, the ON-IN continuum, which seems problematic.

The stimulus artifact problem has an additional dimension, which is that in order to answer the question that Levinson and Meira are asking, you ought to have a stimulus set in which each cell of the etic grid is represented equally frequently. Because what happens here is that, let's say you have a particular type of scene represented much more frequently than another. Let's say you have a very large number of attachment scenes for whatever reason. Then that means those attachment scenes are much more likely to be classified using the same prepositions in language after language, and so they are much more likely to create a cluster in the eventual plot. So part of my criticism here is that some types of notional relations are represented much more frequently in the stimulus set than others, because the stimulus set was never properly balanced for quantitative research. Bowerman and Pederson did not develop this set of pictures for the purposes of a quantitative study like this. So that's obviously very problematic.

Let's now turn to the third point, one that is not unique to this study; it's shared with the next one I want to present. The problem there is that in this type of research, responses across languages are aggregated. So remember Levinson and Meira calculated a similarity matrix for each language and then added them. The similarity matrices across languages, such as let's say this particular pair of pictures here, if that would get 0.2 in one language and 0.7 in

another language, then, you know, the sum of that would be 0.9, and of course what Levinson and Meira did is calculate the average of those similarities across these nine languages.

That means that even if we can overcome the stimulus artifact problem, meaning the extent to which the response is an artifact of the composition of the set of stimuli, we'll still get a picture of how the stimulus set was categorized when we lumped the responses from the speakers of all these different languages together. This does not tell us very much about what the speakers of the individual languages do, and that in my view is also a problem.

Majid, Boster, and Bowerman did a study similar to that of Levinson and Meira, in this case on the responses collected across a larger sample of languages on the basis of the Cut and Break Clips we talked about earlier. It's a set of 61 video clips featuring mostly scenes of cutting and breaking, but there are a few additional kinds of scenes, which I'll talk about in a moment.

Now they had us code the data, similarly to how Levinson and Meira had the data coded in terms of the topological relators that are shared across pictures, in terms of the verbs that are shared across video clips. So again the question is how similar are the semantic categorizations of the notional space of events of cutting and breaking across languages in terms of the extensions of the verbs in which the various scenes fall.

So two video clips would be treated as having been treated similar to one another by the speakers of a particular language if those speakers used the same verb to describe them. Now, in the study of Levinson and Meira, we had the problem of identifying what exactly we mean when we say the speakers of this language use the same preposition or postposition, because you saw that in the case of Yucatec you have these combinations of prepositions and relational nouns that were treated as instances of the same preposition. In the very same way, in the Majid et al. study, there are a lot of complex predicates. So let's say you have, for example in English, *break*, *break in half*, *break in two*, *break up*, and so on and so forth. So those are different complex predicates, all of which are based on the same verb root *break*. So the question is, should we treat those as five different verbs or should we treat them as instances of the same verb? Obviously this is going to affect the outcome of the analysis greatly.

Majid et al. decided to treat all of these as instances of the same verb. This is a decision that you have to make, and as some of you can probably imagine, the reason why they made the decision this way is, if you go with a much larger set of categories, you'll therefore make the probability of two pictures or two clips in this case having been treated the same way in a given language much smaller, which means you are reducing the possibility of finding much

structure of interest in a similarity space. On the other hand, if you go about it this way, you are increasing the possibility, the probability, of finding artificial structure that isn't really in the actual semantic categorization of the speaker.

There's an important difference in the way Majid and colleagues calculated the similarity scores for each language. They did not use averages across the speakers of each language. Rather, they treated a pair of clips as having been treated similarly by the speakers of a given language, if at least one speaker has used the same verb for this pair of clips. So that gives us a solution that in some sense I prefer over Levinson and Meira's approach, because the use of the average strikes me as more problematic. On the other hand, it is still the case that there's no accounting for the fact that in some cases, you know, all of the speakers of a language use a particular verb for a particular pair of clips, and in some languages in some cases only one speaker does. Those two cases are still treated as identical in this approach. So once again the problem of inter-speaker variation, variation across the speakers of the language, is fundamentally unsolved.

A composite similarity matrix from the responses across languages is created, and once again multivariate statistical analysis is performed on this resultant similarity matrix. Although in this case they didn't use multi-dimensional scaling, but rather a series of other techniques: a correspondence analysis, a factor analysis, and a cluster analysis.

Remember I said that when Levinson and Meira created this multi-dimensional scaling plot, they forced the spatial model of the similarity matrix into a two-dimensional space, thereby losing whatever information it cost them to effect this compression, if you will. What correspondence analysis allows you to do is it gives you a measure of the amount of information that you are losing, depending on the number of dimensions that you are reducing the similarity matrix to, or the number of dimensions that constitute the spatial model into which you are forcing the similarity matrix. Each data point represents a dimension of the possible spatial model, and it turns out that the first seven dimensions that show up in this correspondence analysis account for about 62% of the information in the data set, or 62% of the 'variance' in the data set, as the statisticians say.

Now we can then go on and ask, what are these first dimensions? It turns out that the first two dimensions both distinguish between clips that don't actually show events of cutting and breaking, but rather other kinds of events, which I did say in the beginning were included as well in the set of clips when we designed it. Those were events of opening and closing things, such as opening and closing a bottle, or a box, or book, or something like that, and one event where somebody moves a chair away from a table.

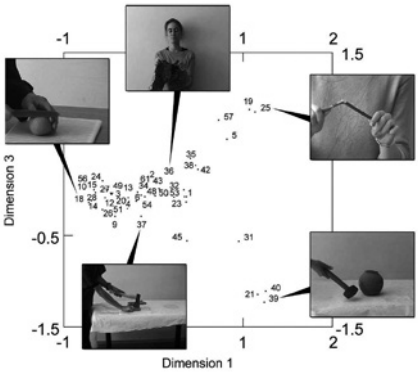

FIGURE 8.1
Plot of the first against the third dimension of the similarity space of the "irreversible" clips
REPRINTED FROM *COGNITION*, 109/2, ASIFA MAJID, JAMES S. BOSTER, MELISSA BOWERMAN, "THE CROSS-LINGUISTIC CATEGORIZATION OF EVERYDAY EVENTS: A STUDY OF CUTTING AND BREAKING", PP. 235-250, COPYRIGHT 2008, WITH PERMISSION FROM ELSEVIER

And why did we include these scenes? Because Melissa Bowerman had observed that children across languages often in the beginning treat the scenes of opening and closing the same way as scenes of cutting and breaking. So for example, they would refer to an event of cutting an orange as opening the orange or something like that. Melissa simply wanted to see whether there are adult languages spoken somewhere in the world that conflate these two types of actions as well.

What we can do is we can simply eliminate these first two dimensions from the data set, and do a new correspondence analysis on the remaining dimensions. And when we do that, we get a new set of dimensions, and we can now do a two-dimensional plot of the similarity space. There are dimensions one and three of the second analysis. So this is basically a two-dimensional plot that was created very similarly to how the MDS plot in Levinson and Meira's study was created. Every point corresponds to one of the video clips, and what this shows us is how similarly the clips were treated when we lump the responses from across these 20-odd languages.

If you have a cluster of clips, that means all of these clips were treated in ways that were very similar to one another, whereas if you have clips that are farther apart, then those were clips that, you know, the speakers of most languages distinguished from the clips in the cluster. Then we can do something that Levinson and Meira did not do, which is we can try to interpret these dimensions, which Majid and Co. did. They argued that the first dimension represents what they call the predictability of the locus of separations.

So what do they mean by that? If I take a stick and I snap it between my hands, I can pretty much predict where the thing is going to break. It's going to be more or less precisely in the middle between my hands. On the other hand, if I take a vase and I swing it over my head and smash it against the

table, then you know there's no particular point at which this thing is going to break. There's no line, no particular region, the whole thing is just going to go to waste, and it's going to turn into small shards. So, the first one would be an example when there is a predictable locus of separation. The second one would be one where there is no predictable locus of separation. That's what they mean.

The third dimension would be one that separates the scenes into a snapping cluster and a smashing cluster, and you can see that you have mostly snapping scenes, and down here you have mostly smashing scenes, and supposedly all other clips somehow submit to this analysis, although I don't really see how.

What's interesting about this locus of separation analysis for the first dimension is the contributors to this study all tried to come up with descriptions of the semantic systems of the verbs in their particular languages, and a number of these descriptions were published in a special issue of the journal *Cognitive Linguistics* in 2007. Nobody actually reported to have found a semantic category of predictability of locus of separation in any of these verbs in the languages that were involved in the study.

So here we have something falling out from an analysis over an aggregate of the categorization data from across the languages, which does not emerge in any actual semantic analysis of any particular verbs that came up in response to the clips. What does that tell us? What is the status of this dimension that seems to structure the similarity space? That's a question I'm troubled by. I'm not sure I know the answer, but I'm suspicious that what we're seeing here may be an artifact of the aggregation procedure rather than something that represents anything about how these clips are categorized in the minds of the speakers of these various languages here.

They did another analysis, which represents the similarity between each individual language's similarity matrix and the aggregate similarity matrix. So this shows you how well the similarity space of the speakers of each language fits the aggregate similarity space. And so you can see which languages are farthest away from the mainstream: Dutch, Swedish, and German—so all the Germanic languages are outliers.

Another way of answering that question is something that we explored in the MesoSpace project with respect to reference frame data from a bunch of different Mesoamerican languages. This is based on data from the Ball and Chair task. So how is this semantic categorization? Well in the sense that there is the same universe of 48 pictures of balls and chairs, and this universe gets sorted by speakers of different languages in terms of scenes that are treated using geocentric frames of reference, scenes that are treated using relative frames of reference, and so on and so forth.

Of course, our question wasn't simply how similar are these categorizations across languages. Rather, in response to the proposal by Peggy Li and Lila Gleitman, are trying to find out to what extent the use of the frames of reference by the speakers of different languages is driven by their native languages as opposed to by factors such as their level of education and literacy.

In this pilot study that I'm presenting today, I'm just going to talk about the linguistic data. I'm presenting data from six Mesoamerican languages; two non-Mesoamerican indigenous languages, namely, Sumu and Seri; and three varieties of Spanish: Barcelonan, Mexican, and Nicaraguan Spanish. This is 11 linguistic varieties altogether, and we have data from five pairs of speakers per linguistic variety.

We did an analysis of the data not over the items, as Levinson and Meira and Majid et al. did, not over the stimulus items, not over the pictures in this case, the ball and chair pictures, and not how they are treated across languages. Rather, we did an analysis over the speakers, the participants of the task. So we are asking how does the behavior of the participants cluster across languages? What's the best predictor of the frames of reference that a particular person uses? Is the best predictor their native language? Is the best predictor the second language they speak, their use of Spanish as a second language in the case of speakers of ingenious languages of Mexico and Central America? Or is the best predictor their level of the literacy and education?

Of course the nativists such as Li and Gleitman and collaborators would predict that literacy and education are the most powerful factors and that the native language and the use of a second language play no role, or only a role that can be reduced ultimately to the other factors, while the Neo-Whorfians would predict the inverse.

We coded the data for eight different response categories, including various types of reference frames based on our fine-grained classification, but also topological responses, meaning responses that are perspective-free and involve no frame of reference. So that means for each participant we calculated a set of eight frequencies. These frequencies can be interpreted as points in an eight-dimensional space, and the distances between these points can be once again interpreted as similarities across the responses of the participants.

This gives us a similarity matrix, but what we did with the similarity matrix is not one of the kinds of analysis that we've seen already, but rather a phylogenetic analysis, and in particular an analysis using an algorithm called Neighbor-net. And what that creates is this kind of really funky diagram, like a spider web.

This is a type of analysis that was originally developed in evolutionary biology for the purpose of statistical analysis over the similarities across biological species. Imagine you go to an island, and you discover 32 different species of

finches on this island, and you want to know what is the most likely evolutionary tree that captures the relations of how these finches have evolved, how in other words are they related to one another. So what you do is you compute a similarity matrix on the basis of the features that the various species and varieties share, the form of the beak, the color of the plumage, you know, their diet and so on and so forth, you get the general idea. And then out comes something like a rootless tree, because you don't want to presuppose a particular ancestor, because you don't actually know what the common ancestor of these species is.

The reason this approach is useful for linguistic typology, and not just for semantic typology but also, for syntactic typology—where by the way it has been pioneered by people like Michael Cysouw and Balthasar Bickel over recent years—is because this gives you a relatively stress-free representation of the similarity space. So unlike a multi-dimensional scaling plot, it does not force you to compress the similarity space into a model of just two or however many dimensions. This is in a sense an a-dimensional model of the similarity space.

This has, as it turns out, a structure where you find most of the participants who use relative frames—so the terminal nodes at the ends of this web are, basically, the participants. They are properly speaking directors of dyads, or pairs of speakers. So in a sense every terminal node stands for a pair of speakers, but it's only the utterances of the director that were actually coded.

It happens to be the case that the shape of this plot is determined by the people with high usage of relative frames of reference falling on the left side of plot and the people with high usage of geocentric frames of reference falling on the right side. Now that's not a foregone conclusion. There are altogether eight different response variables that were included in the analysis, and so it happens to be the case that these two are the ones in which the participants differentiated themselves most. But that's the empirical finding. That was not a foregone conclusion.

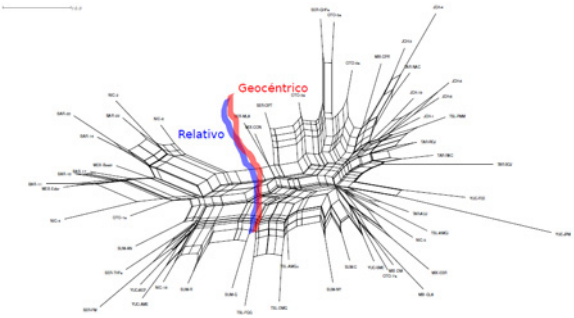

FIGURE 8.2 The Neighbor-net and its "geography"

Secondly, in terms of the second language, if you are looking at the relative pole of the net, you find that almost everybody, with two exceptions, speaks Spanish frequently either as a native language or as a second language, whereas on the geocentric side of the net, there is no clear tendency. You find all sorts of stuff going on. In terms of education and literacy, there are also pretty clear differences. So up to this point, both the expectations of the Neo-Whorfians and the expectations of the nativists have been confirmed. That's not surprising of course, because there is a correlation here in a sense that, generally speaking, unfortunately, is the case that levels of education or literacy are somewhat lower across many, though not all, indigenous populations.

So how can we find out which of the factors are the most powerful? One thing we can do is calculate for each value of each variable the total distances across the responses that share that particular value. Of course, where the sum of these distances is the smallest, you have found the variable that causes the participants to cluster the most. According to the predictions of the Neo-Whorfians, those ought to be the linguistic variables, whereas according to the predictions of the nativists those ought to be the nonlinguistic variables.

It turns out the densest clusters are by the speakers of Spanish. These are all the people who use Spanish as either first or second language. In third respect, it's the speakers of the non-Mesoamerican indigenous languages Seri and Sumu that cluster together, and only in fourth place does a nonlinguistic variable come in, and that's level of education. So, all those people that have no formal education also cluster pretty strongly in terms of their responses. So you see that the three strongest predictors of how people perform in this task are linguistic variables. So they are not nonlinguistic variables.

Now, we've taken this as a preliminary confirmation of the Neo-Whorfian view, but we're only getting started. So in the meantime we're working on a second analysis, a linear regression analysis which will allow us to find out which of the factors actually contributes significantly and whether it is possible to reduce the role of language to a cumulative effect of the nonlinguistic factors as Li and Gleitman propose, or whether the effect of languages is irreducible.

Plus we're trying to do the analysis not just on the linguistic data but also on the nonlinguistic data, the recall memory data. Most importantly, perhaps especially in terms of the headache it has been causing us, this analysis, as far as the demographic data of the participants is concerned, is really only based on the researchers' estimates. So it's based on the researchers' estimates or knowledge of the level of education of the participants rather than the participants' own estimates. So we are trying to collect the participants' own estimates now and do the analysis on the basis of that as well. So this is very much work in progress, and of course, if we're finding what we're looking for, then

we'll repeat the analysis with the larger data set that also includes a ton of non-Mesoamerican languages including Mandarin, Vietnamese, Japanese, and so on and so forth.

And before I leave this topic, I want to say, I used to think we were the first ones to do the analysis over participants rather than over stimulus items, and thereby potentially solve the aggregation problem. Meaning, we allow language actually to play a role as a predictor variable, so we can see the impact of different languages on the categorization of the data directly in a way that the analyses in Levinson and Meira and Majid et al.—since they lump the data from across the speakers of different languages—don't allow us to see.

But actually it turns out that Majid and colleagues did a study on children's categorizations of events of cutting and breaking, where they did do an analysis over participants rather than stimulus items as well, a very beautiful analysis that they haven't published yet. I don't know why. But that actually is quite intriguing, because it shows you that the kids already at very young age behave language-specifically. So you see that all the Dutch kids cluster with one another, and all the English kids cluster with one another, and Korean kids cluster with one another.

Are we ready to go quantitative? Fundamentally no, because we are not ready to go quantitative until we've actually compiled data from the speakers of each language that can be representative of the language community as a whole. That means we have to solve both the intra-speaker variation problem and the inter-speaker variation problem.

Now in the humanities tradition of linguistics, that's something we've never attempted to do. Outside sociolinguistics and to some extent psycholinguistics, that's something we've never attempted to do. We've simply idealized to the native speaker. We have always assumed that, oh, you know, either I'm the speaker myself, so I'm just going to go with my own intuition, I'm going to do a paper on my own intuitions. Or else I'm not a native speaker and I'm going to find myself a native speaker, I'm going to do a paper on his or her intuitions and that's my database.

In quantitative research, that's just not a legitimate move. If we want to apply quantitative methods to the study of crosslinguistic semantic categorization, then we have to be serious in how we sample from speakers of different languages, and we have to make sure that we're doing it in a way that is actually valid by the standards of quantitative research, which so far hasn't happened. This in a sense means that all the studies that we've looked at today, probably will still have to be considered sort of pilot studies. So in the future we'll get hopefully better designs with methodologies and protocols that are more respectable by the standards of the social and behavioral sciences.

In very quick summary, typology is inherently a quantitative research program in that it seeks to formulate generalizations over the distribution of linguistic properties across the languages of the world, and that's by nature quantitative data. Statistical methods have been employed in syntactic and phonological typology since the 1980s. In semantic typology, statistical analyses have been developed during the last decade.

The nature of the data that forms the input to statistical analyses typically differs across morphosyntactic, phonological, and semantic typology. In morphosyntactic and phonological typology, typically data from large language samples are analyzed, but at a single data point per language completely abstracting from inter-speaker variation. See Mathew Dryer's database of around 1800 languages underlying his statistical tests of the Greenbergian word order typology. In contrast, semantic typologists draw on much smaller language samples. The largest one that we've seen today was Majid et al.'s sample of 20-odd languages in the Cut and Break study. This is based on primary data, so it gives us the ability to take into account inter-speaker variation.

The early studies largely ignored this option, choosing to aggregate the responses of the participants from each language community in some form or another. This was the case both in Levinson and Meira's and in Majid et al.'s study. One alternative is to do what Regier and colleagues did, namely, to compute some population-internal measure of the semantic categorizations within each language community. But that has its own problems, namely, the question then of course is where do we get this supposedly language-independent measure of naturalness from.

So in the case of Regier et al., they used a model of the visual color space, which has been assumed to be universally valid, even though it's apparently based originally on the responses from just one French speaker. Then we looked at the approach of doing a multivariate analysis, not over the similarity space of the items, but rather over the similarity space of the participants. That allows us to treat language as one of the predictor variables and to directly measure the effect of language as opposed to other factors on the categorization of the items.

So far there are several problems that have been neglected, due to, you know, the influence of the humanities traditions in linguistics, such as, there are so far no clear criteria to ensure the validity of the samples that we've been using in crosslinguistic research, and there is no clear approach to dealing with intra-speaker variation.

LECTURE 9

Event Description: Variation at the Syntax-Semantics Interface

Now I'd like to look at semantic typology in an area that we haven't looked at much, although we've touched on it, and that's the semantic typology of event representations in language. That has come up at least in one study before, namely, the study on the conceptualization of semantic categorization of events of cutting or breaking. I'd like to ask some more basic questions about how events are represented in language and across languages.

I'd like to start with a problem that's been bugging me pretty much as long as I've studied linguistics. I've long been interested in the phenomenon of serial verb constructions or multi-verb constructions, which you find in languages all over the world, including, on some accounts, in Mandarin and other Chinese languages. One distinction that you very commonly find in the literature is between constructions—serial verb constructions and others—that represent an event as a single event versus a sequence or a combination of multiple events.

I have some quotes for you to illustrate this intuition, starting with an observation by Mark Baker, who was writing about serial verb constructions in African languages in particular. He's saying "(…) true SVC structures and covert coordination structures seem to feel different to native speakers." "Feel different to native speakers," that's one heck of a scientific statement right there. "The covert coordination tends to be perceived as a sequence of distinct

All original audio-recordings and other supplementary material, such as any hand-outs and powerpoint presentations for the lecture series, have been made available online and are referenced via unique DOI numbers on the website www.figshare.com. They may be accessed via this QR code and the following dynamic link: https://doi.org/10.6084/m9.figshare.11419149

© JÜRGEN BOHNEMEYER, 2021 | DOI:10.1163/9789004362628_010

events, whereas the SVC is perceived as a single event (…)" [Baker 1989: 547][1] Similarly, here's Bob Dixon: "A SVC consists of more than one verb, but the SVC is conceived of as describing a single action." [Dixon 2006: 339][2]

The question I have is just simply what does that mean? I can give you a description of this entire lecture that's as simple as possible. As a matter of fact, I just did. I called it a *lecture*. You can't come up with a more simple description than that. On the other hand, I could also say, well, the chairman of the department gave an introduction, and then I started with a very round-about introduction to the topic, trying to situate it, and then I gave people a few quotes, and so on and so forth. Now what am I doing? Am I describing the same action or not? The description is clearly more complex. But isn't it still a description of the same lecture and isn't the lecture still an action? So is it a simple action or is it a complex action? I don't know. Here is one by Sebba from a classic volume on serial verb constructions.

> Although two or more verbs are present, the sentence is interpreted as referring to a single action rather than a series of related actions. Although the action may involve several different motions there is no possibility of a temporal break between these and they cannot be performed, for example, with different purposes in mind.
>
> SEBBA 1987: 112[3]

The action cannot be performed with different purposes in mind, whatever that means. So here is the question, what does this mean? What is a single action? What is a single event? And how do we know that a linguistic expression is a description of a single event?

Let's take a step back and try to understand why it is that we're having this conundrum. When we're talking about objects, objects consist of parts, and these parts are located in space together. And what holds them together is what the Gestalt psychologists used to call the 'Law of Common Fate.' That is, if you move one part, you have to move all the others, or else you have to break up the object.

Obviously, that's not the way it works for events. The parts of events are in some sense located together in time. But it's much less clear what exactly it

1 Baker, M. C. 1989. Object sharing and projection in serial verb constructions. *Linguistic Inquiry* 20: 513–543.

2 Dixon, R. M. W. 2006. Serial verb constructions: Conspectus and coda. *Serial verb constructions: A cross-linguistic typology*, ed. by Alexandra Y. Aikhenvald and Robert M. W. Dixon, 338–350. Oxford: Oxford University Press.

3 Sebba, Mark. 1987. *The syntax of serial verbs*. Amsterdam: John Benjamins.

is that holds the subevents together. In other words, under what conditions would we think of a sequence of two events in terms of forming a superordinate whole the way we think of the parts of a shelf forming parts of a single object.

So the problem is in short that of the upper bounds in mereology, in other words in part-whole structures. The sentence that I'm producing right now is part of this lecture. The lecture is a part of a series of ten lectures. These ten lectures are part of my visit to China. They are also a part of what I'm doing this summer. They are also a part of what you're doing right now, of course. But anyway, ultimately all of this is part of the entire history of the universe. And why can't we think of the history of the universe from its beginning, if it had one, to its eventual end, if it has one, as a single event? So the problem is you can conceptualize events as being part of superordinate events, and there is no obvious intuitive upper bound to these part-whole structures until you've reached the largest one we can think of, which is the history of the universe.

Obviously you can describe one and the same event in a way that makes the part-whole structure visible, or you can describe it abstracting away from that internal structure. So, for example, suppose Floyd, our imaginary hero, takes a trip from Nijmegen in the Netherlands to Düsseldorf in Germany. You can drive there across the border in a little over an hour, or you can take a train. It's a nice trip.

Anyway, so we can say, if we are speakers of English, [(11)]:

(11) *Floyd went from Nijmegen to Düsseldorf via Moers*

And this is a single-clause description that employs a single verb. I can also break it down across two sentences, saying (12):

(12) *Floyd left Nijmegen. He passed through Moers and then reached Düsseldorf.*

So that's three verbs, two sentences. Or I can say (13):

(13) *Floyd went from Nijmegen to Düsseldorf, passing through Moers on the way.*

The intuition that comes with these different descriptions is that each presents what is the same event in extra-linguistic reality—or at least what we think of as the same event in extra-linguistic reality—at a different level of granularity and at a different level of making the internal complexity visible.

So going back to these initial statements here, these initial quotes, our task is to compare structures or constructions of event descriptions across languages here as a semantic typologist. We want to know under what conditions we can say that an event description is simple. Under what conditions can we say that an event description describes a single event? We want some way of operationalizing this intuition such that we can actually test it.

How do we do that? Since we're actually interested in comparing constructions in different languages in terms of how much information they convey about the cognitive event representation in the mind of the speaker, we can't just use the syntactic properties of the construction itself. We can't just say one language uses a combination of three verb phrases and another language uses a single verb phrase to describe this event. The first problem with that is we don't know what is a verb phrase in one language corresponds to what's a verb phrase in another language. Second problem, in a sense, we're precisely looking for a way of defining event description constructions language-specifically such that the semantic property of describing an event as a single whole versus as a complex structure enters the description of this construction. So if we use the structural properties of the construction itself as a defining criterion that would make it potentially circular.

What we're looking for is a measure of event segmentation that is sensitive to the syntax of the event descriptions across languages, but that's applicable across languages independently of the particular language-specific construction type. The solution that I've been proposing to this problem is what I've called the Macro-Event Property, with a label that I'm borrowing from Len Talmy's work. But, whereas Len Talmy has been working on macro events as a type of conceptual representation, I'm interested in this Macro-Event Property as a property of event description constructions, a form-to-meaning mapping property, a property that assesses how much information about an event, or a conceptual representation of an event, a particular construction conveys.

The way it works is by tapping into what it is a native speaker can say with this particular construction about an event they're conveying information about, an event they are describing. The idea is specifically to make use of properties that ontologically define events as events and separate them from other ontological categories—in particular, the fact that they have a location in time.

We can tap into this property, but we're going to tap into it from the perspective of how much information a given construction can convey about the location in time. We're going to do that specifically by looking at time positional modifiers—time adverbials such as *at eleven*, *on Wednesday*, *at noon*, *in the morning*, and so on and so forth.

If we go back to the description that we started out with, it turns out that at this level of description, we can assign each of these clauses or verb phrases a separate time-positional modifier. That gives us (14):

(14) *Floyd left Nijmegen at 11:00am. He passed through Moers at noon and reached Düsseldorf at 12:30pm.*

This means in this description we can make the location in time of three subevents explicit. And in that sense we can say that the description allows us to distinguish among three different subevents.

In contrast, if we choose the single-verb description (11), which only expresses the different subevents in terms of path relations reflected by the prepositions that occur in this motion event description, then we aren't able to add those separate time-positional modifiers. So (15) is clearly ill-formed:

(15) ?*Floyd went from Nijmegen at 11:00 to Düsseldorf at 12:30 via Moers at noon*

This is interesting because it's very obvious what the speaker was trying to say here. And nevertheless, English speakers have the feeling that this is too much of a good thing. Having these three time positional modifiers here is too much information. It's as if, you know, as people say, three is a crowd.

What we can do is have a single time adverbial that has scope over all three subevents, situating them in time together, as in (16):

(16) *On Wednesday, Floyd went from Nijmegen to Düsseldorf via Moers.*

Notice that now we're treating these three subevents as parts of a single event. And that's precisely what we've been looking for. Now we found the beginnings of a way to make explicit this intuition that in this case (16), we're presenting this journey, this motion event as a single event, whereas in (14), we're representing it as a sequence of three events. We found the beginnings of a way of making that intuition explicit. We can say that the sense in which we're presenting this motion event as a sequence of three subevents in (14), namely, a departure event, a path event and an arrival event, is the sense that this description (14) allows us to assign each of these events a separate position in time, whereas in this case (16), we're talking about the same three subevents, but we're no longer treating them as separate events. That's apparent from the fact that we cannot locate these three events in time separately. So the only thing that we can locate in time now is the entire complex superordinate event. In that sense we're talking about the three subevents of this trip as forming a single event together.

So what we're aiming for here is a mapping property of constructions. It's a property of constructions that determines the possible semantic properties of these constructions. It's not an ontological property. It's not a semantic category or ontological category of macro-events. That's important to understand. I'm not saying that there are macro-events in nature. I'm not saying that the event we're talking about is a different event in (14) with respect to (16). On the contrary, I'm saying that for all we know these are two different descriptions of exactly the same event.

Now it may be the case that the underlying conceptual representation in the mind of the speaker is different. In fact, at the level of thinking-for-speaking, at the level of forming an underlying conceptual cognitive representation for the purposes of translating it into a verbal utterance, that has necessarily to be the case. Let me put it in this way: the type of cognitive representations that Len Talmy talks about as 'macro-events' may or may not map into this Macro-Event Property at the syntax-semantics interface. That's something I'm not sure about. So for now I want to make sure that we understand that the Macro-Event Property is a mapping property that describes the amount of information that particular descriptions convey about the events they describe and nothing more.

But here is the new definition itself: "A construction C that encodes a (Neo-) Davidsonian event description $P(e)$…"—Donald Davidson invented a formal language for describing the logical properties of event descriptions in the 1960s, a kind of predicate calculus for event descriptions, if you will. [Continuing:]

> A construction that encodes an event description has the MEP if that construction has no constituent—*syntactic* constituent, that is—that describes a proper subevent of the event such that that constituent is compatible with time-positional modifiers that locate the runtime of the subevent, but not that of the larger event.

It sounds complicated, but let me explain why I'm talking about constituents that describe subevents, which is the way that this definition differs from the old one. So consider this example: *On Monday, Sally read the letter that Floyd had written on Sunday*. Obviously we have two different time positional modifiers here, and they talk about different events. So that would seem to suggest that this construction doesn't have the Macro-Event Property. In other words, it's a construction that allows us to talk about separate subevents that each have their own location in time, and each we can assign separate locations in time to.

However, on closer inspection, we notice that because the second event is described by a relative clause, it's not actually a part of the first event. It's not the case that Floyd's writing event is a part of Sally's reading event. Intuitively this makes sense, because the relative clause here is a noun phrase modifier. It modifies the letter. It does not actually contribute directly to the event description. So we would not want to say that this overall description here lacks the Macro-Event Property just because it contains a relative clause, which doesn't serve to describe a subevent, but merely to identify a participant involved in the event described in the matrix clause.

Furthermore, there is a second problem that's long been bugging me. In English, at least according to some speakers, you can say stuff like *Floyd drove from Rochester to Buffalo at noon*. It takes about an hour and ten minutes to drive from Rochester to Buffalo. You can take my word for it, because I do it three times a week. That is, if there's no blizzard. If there is a blizzard, then all bets are off.

So what does it mean to say this happened at noon if it takes an hour and ten minutes? You can understand this sentence such that *at noon* doesn't actually refer to the location of the entire event, but merely to the time of the departure. So what this really means is *Floyd departed from Rochester, headed for Buffalo*. As a matter of fact, I believe when people use this time adverbial in this way, they interpret *to Buffalo* not actually as a goal phrase that marks the location where Floyd arrived, but rather as a directional specification. It's something like that Floyd departed from Rochester at noon, headed in the direction of Buffalo.

However that works, what's important here is that *at noon*, under this interpretation, refers to a subevent. Since *at noon* in this case is still a modifier of the clause or verb phrase as a whole—the description does not contain a constituent that refers to the departure only, or conversely, if there is such a constituent, it would be this prepositional phrase *from Rochester*, but there is no syntactic relation between this prepositional phrase and the modifier *at noon*. Therefore, once again we can treat this description as having the Macro-Event Property despite the possibility of construing this modifier as referring to a particular subevent only.

That was the introduction to the Macro-Event Property. Now what I would like to do for the rest of this lecture is present three case studies that employ this property in typological research. The first one is the one that we reported in that 2007 article that first tried to define the Macro-Event Property formally. This is a case study on the segmentation of motion event descriptions. The second one concerns the segmentation of causal chains, so the domain of causal

relations in language. The third one is this paper I've been working on with my colleague Robert Van Valin on the syntactic correlates of the Macro-Event Property across languages. In other words, the question we're asking is, is there something like a macro-event phrase in the languages of the world? And we have a positive answer to that question.

Let's take a quick look at the methods and the data that that we've been drawing on in this study. The idea has been to identify a particular domain of complex events such as the motion domain or the domain of causal relations; to identify constructions that across languages are used to encode or describe events in this domain that have the Macro-Event Property; and then to ask what further mapping properties these constructions have in common aside from the Macro-Event Property.

In the domain of complex motion events, meaning events that involve sequences of location changes of a single figure with respect to multiple grounds, our interest—this was the first case study we did on events segmentation, the structure of linguistic event representations. We started with motion because of the presumed universality and the presumed conceptual simplicity of the motion domain, which psychologists and philosophers have long insisted on, but also because of the strong perceptual bias of motion events in the sense that a motion path gives you a spatial trace of the temporal history of the event, and you can use the spatial trace to map the segmentation of the event into subevents in space.

So then we've looked at causal chains. The data that we have been drawing on for both of these studies were collected using the same set of tools initially, although we then developed an additional stimulus set for the study of causal chains for reasons that I'll explain in a second. So these tools, which I'll introduce in a moment, have been applied to this set of languages by these researchers, whose contribution I gratefully acknowledge, and they generally tested three to five speakers per language.

The stimuli they used consist of basically two parts, plus the third that's involved in the causality study only. The two parts are a set of animations, short animated video clips, featuring events of the relevant kind—complex motion events, complex causal chains, and also complex transfer events, meaning change of possession events, which we haven't actually looked at, meaning we never looked at the data that were collected with those.

In addition, I also devised a questionnaire, which is supposed to define an etic grid for this semantic typology in such a way that the researchers can use that grid to make sure they have as many examples as possible of all the cells of that grid, whether they get it through the descriptions of the video clips or any other way. Maybe they find examples in narratives they collected or

conversational data or from other elicitation sessions or maybe they come up, if they can't find any instances in their data, with ways of eliciting these particular cells themselves.

What we tried to do is to elicit both preferred descriptions of all of these stimulus items and also ranges of possible descriptions for each language. We conducted tests to ensure that all the subevents represented in each stimulus scenario are linguistically encoded in the responses and obviously we used the time adverbial tests to establish whether the responses have the Macro-Event Property or not. And we used other tests as well.

In addition, for the study of causal chains, for reasons that will become apparent once I present the results, we complemented the ECOM clips—ECOM by the way stands for 'event complexity.' So there's an initial series of 74 video clips. That turned out to be not sufficient for getting rich material on causal chains, not actually so much because of the nature of the videos but rather because of the nature of the tasks that we gave the participants, the speakers of the languages we worked with.

We used an additional task, which we called 'ECOM Causality Revisited,' precisely because we had to revisit the question of the encoding of causality after the initial task didn't quite produce useful results. I'll explain when I get there why that is. What the new task did is ask questions such as 'Why did it happen?', 'Why did the triangle break?', 'Who broke the triangle?', 'Who caused the triangle to break?'—which you can think of as analogs to the Where-elicitation prompt in the Topological Relations Pictures Series, the BowPed pictures, which we've talked about several times during this lecture series.

With respect to the BowPed pictures, the researcher asks the participants, 'Where is the figure?' The participants' task is to answer the question 'Where is the figure?' In this approach, the participants' task is to answer questions such as, 'Why did the triangle break?' And 'Who broke the triangle?', 'Who caused the triangle to break?', and so on.

Now let's take a look at the findings, first of all, in the motion domain. So let me talk a little bit more about the etic grid that underlies the construction of the stimuli and the elicitation that we did. So this is actually very similar to Herb Clark's work from the early 70s on the conflation of those topological relations AT, ON, and IN in locative versus motion prepositions in English.

This is the manifestation of the same ideas in Ray Jackendoff's Conceptual Semantics framework. The idea is there's a set of primitive path functions which describe the source, the goal, and the route of a motion event and in addition the directions toward or away from some reference entity. These correspond to subevents of departure, arrival, and passing, whereas the directional phrases can correspond to any phase of a motion event that's oriented in a frame of

reference. And Jackendoff calls path specifications that involve FROM or TO path functions 'bounded paths'. Examples would be *from the entrance*, *off the roof*, *out of the kitchen*, *to the entrance*, *onto the roof*, and *into the kitchen*. 'Route' descriptions, the ones that describe passing events and involve VIA functions, would be exemplified by *past the entrance*, *across the roof*, or *over the roof* and *through the kitchen*. And examples of directional descriptions would be *towards the entrance*, *northbound* in the sense of *The train went north*, *down* as in *The plane went down*, *upriver* as in *The detachment of soldiers marched upriver* or something like that, and *left* as in *Floyd exited stage left*. So this gives us a sense of the conceptual components of motion paths and the corresponding subevents that we can use to predict the possible conceptual complexity of motion event descriptions, in other words, to construct an etic grid.

Now for an understanding of the results we got, we need to know a little bit about Len Talmy's typology of motion event framing, which I'm assuming many of you are familiar with. But for a brief recap, satellite-framed languages are languages that present the path information outside the main verb root, and they can do that using a verb prefix, or for example a preposition, or a case marker, or they can use a combination of these things as in this Russian example, *Šarik vy-katilsja iz karobki* 'the ball rolled out of the box.' So in a broad sense, all of these represent the path outside the main verb root, which in this case refers to the action of rolling. In a narrow sense, Talmy would consider only the preverb, the verbal prefix, a true satellite. But terminology aside, the preposition and the case marker also convey information about this particular path function, the source path function.

On the other hand, verb-framed languages are languages that convey path information, information about the path functions in Jackendoff's sense, in the main root. And here is a Yucatec example, a description of the same scenario:

(17) *Le=bòola=o'* *h-hóok'* *ich* *le=kàaha=o'*
DET=ball=D2 PRV-exit(B.3.SG) in DET=box=D2
'The ball, it exited (lit. in) the box.'

So the source path, the OUT OF path function, which gives you a source with respect to a containment relation—in other words, a source in a container, so a motion out of the container—is expressed in the main verb root here. The preposition in this case actually does not convey any path information, because that preposition can also be used for 'The ball went into the container' or even for 'The ball is in the container.' So this preposition is completely path-neutral.

It's also possible to convey information about the path function *both* in the main verb root and in some expressions outside the verb root. This is the case in many of those languages that traditionally have been considered the best examples of verb-framing, such as Spanish. So in the Spanish representation *La pelota salió de la caja* ['The ball exited from the box'], you actually have the source path reflected, not just in the verb *salió* 'exit', but also in the preposition, which is an ablative preposition. So I'm going to argue that strictly speaking this is more than verb-framing, in the case of Spanish it's actually something like double-marking, whereas an example of pure verb-framing is what you get in Yucatec.

What we found is that the languages of our sample fall into three types with respect to the segmentation of motion events. The first one allows presenting all three subevents of a conceptually maximal motion event according to our etic grid—namely the departure event, the passing event, and the arrival event—to present the combination of these three subevents in a single macro-event expression. In other words, in expressions that have the Macro-Event Property. So this is possible in Dutch, but it's also possible in Ewe, which is a serializing language, and similarly in Lao. It's possible in Marquesan, which is a Polynesian language spoken in French Tahiti, and in Tiriyo, which is a Caribbean language.

The second type gives us in a sense a variation on the first type. So it is possible in languages of the second type to describe all three subevents in a single macro-event expression. But that depends, especially if we're dealing with a passing event. If a passing subevent is a part of the conceptual representation, then it depends on the type of the passing event whether a single macro-event expression is possible or whether we have to break this down into a combination of two macro-event expressions. And that type is instantiated for example by the Central Australia language Arrernte and by Basque, Hindi, Japanese, and Trumai, which is an isolated Amazonian language of Brazil.

Finally, Type-III languages are languages that require a separate macro-event expression for each of the three subevents. So these are the languages where you just cannot say 'He went from A via B to C' or '[He went] from A to C via B'. You have to say 'He left A, and he passed B, and then he arrived at C'. So that's the case in Yucatec, and also in Zapotec and Tseltal, languages spoken in the same area. But it's also the case in the Mande language Jalonke; in the Oceanic language Kilivila; in Saliba, another Australian language, in fact another Oceanic language; in Tidore, which is a Papuan language spoken in Indonesia, and Yeli Dnye, the language Steve Levinson has been working on, which you remember from our discussion of color terminologies. All of these

languages, and to be sure many more around the world, actually require three different macro event expressions to represent linguistically what from the perspective of English seems a quite simple event.

Type-I languages are satellite-framed languages, and satellite-framed languages that we've looked at are basically all of this type. So satellite-framing is a very powerful strategy. It allows you to roll multiple subevents into a single macro-event description in a way verb-framed languages generally can't do. This is illustrated here with an example from Tiriyó. 'The cow is bringing the little stick from the tree along the path to the vehicle.' Obviously in order to do that, you have to have a way of describing the three path functions independently of the event classification expressed in the verb.

Take an example from Ewe, which employs a serial verb construction. You get 'The circle rolls from the blue place on the road passing the side of the house going to the triangle.' This is a construction that involves three verbs, each coming with its own syntactic projection. In other words, three verb phrases or three verbal cores, whatever you want to call them. But these together form a single clause, and more importantly you can describe this same scenario using a more complex serial verb construction, which employs directional particles, and these directional particles function as the glue that holds these verb phrases together.

I'm going to say this is not a formal description. This is not a syntactic analysis, calling them 'glue.' What's important is that they are not actually coordinating conjunctions. My former colleague and collaborator Felix Ameka demonstrated in his dissertation that these guys are not coordinators. But crucially, if you use this more complex serial verb construction, then you can have separate time-positional adverbials in each verb phrase. So that means this description doesn't have the Macro-Event Property.

If you go without those directional particles, in other words if you go with the simple serial verb construction, then the only possible time-positional modifier is one that locates all three subevents together. In other words, you cannot individuate the temporal position of each subevent, using an adverbial that has scope over just one of these subevents, and in that sense, this description does have the Macro-Event Property.

Type-II languages are verb-framed languages, but they're double-marking, meaning path is encoded, not exclusively in the verb root, but in addition also reflected outside, as we already saw in Spanish. Here is a Japanese example, 'Someone went from the tree to the house.' So you have the path verb here, but you also have a reflection of the path function in these postpositions or case markers. That allows you to package multiple subevents into a single clause

such that that single clause has the Macro-Event Property. So this description allows only time-positional modifiers that have scope over both subevents.

Now it's not normally possible in Japanese to add a description of the route path. If you want a route path, then you're going to have to add a separate macro-event expression, so as in the example 'Leaving the tree (at eight), crossing the river (at four), one arrived at the house (at five).' This lacks the Macro-Event Property but it does allow you to describe all three subevents, including this crossing event, which is a type of passing event, meaning it involves a route path function. But there is an exception.

That's what distinguishes this type from the third type, which is if the beginning and end location of the passing event happened to coincide with the source and goal locations, then you can have all three functions expressed in a single macro-event expression. Example: *John crossed the Bay Bridge from San Francisco to Oakland*. That's possible in a single macro-event description simply because, as many of you know, the Bay Bridge actually connects San Francisco and Oakland. So you know, as soon as you're crossing the Bay Bridge, you are automatically crossing it either from San Francisco to Oakland or from Oakland to San Francisco.

But if you wanted to say that John went from Berkeley, crossing the Bay Bridge from Oakland to San Francisco, and then he got to Stanford—which is again an event that has a starting point, an end point, and in between the Bay Bridge is crossed—you cannot do that in a single macro-event expression, because obviously if you say *Floyd crossed the Bay Bridge from Berkeley to Palo Alto*, that would suggest the Bay Bridge connects Berkeley and Palo Alto, which it doesn't.

Finally, Type-III languages are the purely verb-framed languages, such as Yucatec, where you need for each of the subevents, for each of the path functions, a separate verb projection. Since Yucatec and other languages of this type lack the kind of serial verb constructions that allow languages like Ewe and Lao to string multiple of these verbal cores or verb phrases together to form a single macro-event description, therefore you get a separate clause. Each of these clauses is its own macro-event description. So in other words, what distinguishes Ewe and Lao from Type-III languages is the ability to form these serial verb constructions, which Yucatec and Zapotec and many other languages lack.

Let me wrap this case study up by briefly mentioning some properties that we've so far found shared among all the different macro-event expressions that we've found across the languages of the world. So, for example, the unique assignment of semantic roles—that's something that holds for semantic roles

and motion event expressions as well. *The ball rolls from the rock across the tracks to the hills* is fine. But ?*The ball rolls from the rock to the hill to the hole* etc. is problematic, because it assigns the goal path function multiple times. Similarly why is it that it's funny to say ?*Sally walked out of the library from the reception to the entrance*? Apparently, the reason is that the source function is assigned twice here.

Now as soon as you introduce coordination, *Sally walked out of the library and from the reception to the entrance*, the sentence becomes fine. But notice when you do that you also lose the Macro-Event Property. Because now you can say *Sally walked out of the library and then from the reception to the entrance* or *Sally walked out of the library and a moment later from the reception to the entrance*. And similarly you can't say ?*Sally went to Nijmegen home*. You can of course say *Sally went home to Nijmegen*, but then *home* and *Nijmegen* actually refer to the same place. They're in an appositive relationship syntactically. But if you say *Sally went to Nijmegen and then home*, that's perfectly fine, but that's a combination of two macro-event expressions.

It seems that there is a correlation between the Macro-Event Property and this constraint on the unique assignment of semantic roles which was first proposed by Charles Fillmore all the way back in 1968 and which nowadays people know under labels such as the 'Biuniqueness Constraint,' that's John Bresnan's label, or the 'Theta-Criterion,' that's what Chomsky calls it. It seems that this property is sensitive to the Macro-Event Property in a sense that constructions that have the Macro-Event Property are subject to this constraint whereas constructions that lack the Macro-Event Property are not.

So for example, in Ewe, if you remember those two types of serial verb constructions, the plain one, the simple serial verb construction obeys the unique assignment constraint whereas the directional serial verb construction does not. And similarly, Japanese has a bunch of different converb constructions, and those that have the Macro-Event Property seem to obey unique assignment of semantic roles, whereas those that lack the Macro-Event Property also fail to obey unique assignment.

The conceptual connection between the Macro-Event Property and unique assignment is the idea that semantic roles are part of the criteria that individuate events, in the sense that if you're talking about two agents, then either you're talking about two events or you're talking about the two agents being involved in the event collectively. So if you have *Sally and Floyd played piano*, there's two ways of interpreting that. If they are separate agents, that means you're dealing with two events, and under this construal you can say *Sally and Floyd played piano*—let's say—*on Friday and Saturday, respectively* with the understanding that Sally played the piano on Friday and Floyd played it on

Saturday. The other possibility is that you ascribe collective agency to them, such as, you know, Sally and Floyd played four-handed piano or something like that. There is a single assignment of the agent role to Sally and Floyd collectively, and under that construal the description also has the Macro-Event Property.

The underlying phenomenon is whether you say *Sally walked out of the house into the garden* or whether you say *Sally walked into the garden out of the house*, the path you're describing is actually the same. So there is no way of understanding this one to the effect that Sally walked first into the garden and then out of the house. That, too, is a property of macro-event descriptions.

Basically the same ground, the same referential ground, the same reference entity, cannot be referred to more than once within the same macro-event expression. So you can say *Floyd went from the first tree to the second tree*, but it's weird to say *Floyd went from the tree to the tree* or (…) *from the tree to it*.

Similarly it's funny to say, and this is actually how I discovered this, my former colleague Bhuvana Narasimhan pointed this out: why can't you say? *Sally went into the tunnel out*? Meaning 'into the tunnel and out of the tunnel,' 'into the tunnel and then out of the tunnel,' which is in fact fine. *Sally went into the tunnel and then exited the tunnel again* is perfectly fine. It's just that that's not a macro-event description. So if it's a macro-event description, then it'll be subject to this binding principle.

Finally there's the Unique Vector Constraint, which talks about the amount of directional path information that can be conveyed in a motion event description that has the Macro-Event Property. So in a scenario where you have a ball rolling inside a container, and then up the wall, and then down on the other side, on the outside of the container, and it rolls onto the triangle here and then up to the top, you have to more or less to break it down into a sequence of macro-event expressions just the way I just did, regardless of the language you're looking at.

If you're describing it in Ewe, you're getting the complex serial verb construction that involves the directional particles. Dutch, which is a satellite-framed language, allows you to conflate the source, the passing event, and the arrival at the goal in a single macro-event expression. But you can't do that with respect to one in Ewe if you want to encode all these different path segments and the directional information that characterizes each of them.

Take a very simple example that illustrates the principle. You can say *The figure moved away from A toward B*, but it would not be an appropriate description of this one to say *The figure moved away from A and then toward B*, because any subevent that involves motion away from A is also a subevent that involves motion toward B. Conversely, where you have directional change involved, you can say *The figure moved away from A and then toward B*, whereas it would be

misleading to say *The figure moved away from A toward B*, because this suggests that all the subevents are events that are both away from A and toward B, which is not the case. In this case there is no single subevent that fulfills both of these specifications.

Now we're going to look at the application of the Macro-Event Property as a diagnostic for the semantic complexity of event descriptions across languages. We're going to apply this approach to the question of which kinds of causal chains get encoded by which kinds of constructions across languages. Initial evidence for typological variation comes from a classic paper by Andy Pawley on the representation of complex events in the Papuan language Kalam, in the highlands of Papua New Guinea. That paper was a big inspiration not just for this study on causal chains, but for the entire project on event segmentation across languages. So in a sense Pawley introduced some of the questions that I tried to answer, coming up with this Macro-Event Property as a criterion.

So Pawley also reports that in Kalam, if you want to say something like 'The rock broke the glass'—which again is one of those apparently simple event descriptions in English. It involves two participants and an apparently simple state change event, so this is a simple enough causal chain apparently from an English perspective. But if you want to describe this in Kalam, you have to use at least three verbs. You have to say 'A stone fell and struck the glass and it broke,' so you have to describe the motion of the rock, the contact between the two objects, and the state change separately.

I mentioned before that in order to get data on the descriptions of causal chains across languages, we had to do a little more than we originally thought we would have to do. So we used those same tools, the ECOM clips and the questionnaire. The ECOM clips and the questionnaire included causal chains among the stimulus scenarios, but the descriptions we collected with that material turned out to not be useful for our proposes, because across languages, independently of the particular language, it turned out that these descriptions were broken down fairly finely across a multitude of macro-event descriptions.

As an example of that, this particular Yucatec speaker says:

(9) *K-u=chíik-pah-al* *le=chan* *kwàadradro=o'*
IMPF-A.3=appear-SPONT-INC DEF=DIM square=D2
chich *u=tàal=e'*, *k-u=koh-ik*
hard(B.3.SG) A.3=come(INC)=TOP IMPF-A.3=collide-INC(B.3.SG)
le=chan (...) *sìirkulo=o',* *le=chan* *sìrkulo* *túun=o'* *k-u,*
DEF=DIM circle=D2 DEF=DIM circle then=D2 IMPF-A.3
óolbèey, *estée,* *k-u=lúubul* *hun-p'éel* *chan*
it.seems HESIT IMPF-A.3=fall-INC one-CL.IN DIM

che'-il *yàan* (...) *ti'=e'*,
wood-REL EXIST(B.3.SG) LOC(B.3.SG)=TOP
ku=hats'-ik *le=chan* *triàangulo=o'*,
IMPF-A.3=hit-INC(B.3.SG) DEF=DIM triangle=D2
k-u=káach-al
IMPF-A.3=break\ACAUS-INC
'(...)the little square appears, it comes on hard, it bumps into the little (...) circle; the circle, uhm, a little stick apparently falls, that it has, it hits the triangle, the triangle breaks'

Notice, for example, that the breaking of the triangle by this yellow rectangle is represented as an uncaused state change event. So *the triangle breaks* is what the speaker says. And to do so, the speaker actually derives a sort of middle voice form, an 'anti-causative' form of the verb, so the inverse of a causative, an anti-causative form that turns a root that talks about caused breaking—meaning this is a transitive verb root—into an uncaused state change description, a state change description that's abstracted away from the cause. Which is very curious—why is it that the speaker goes to the trouble of deriving a simple uncaused state change verb from a transitive stem if the transitive stem would be perfectly suitable for giving a more condensed representation of this causal sub-chain here?

My hypothesis is that the answer is a principle of the distribution of information in discourse, which says you've already mentioned the movement of the yellow bar, so it wouldn't be efficient to mention the yellow bar again in order to present it as the agent of the breaking event or the causer of the breaking event. More importantly, if you do that, it actually triggers an implicature to the effect that the referent is not the same yellow bar. So if I say *The yellow bar hits the triangle and then the yellow bar breaks the triangle*, that's a funny description. It suggests that we are talking about two different yellow bars. I would have to pro-nominalize it—*The yellow bar hits the triangle and it breaks it*. But that's still intuitively an awkward, and clumsy description. What we're getting instead is a description that explicates, if you will, each subevent, but leaves the causal links between the subevents implicit. In other words, the causal links between the bumping and the dropping and the hitting and the breaking are left to stereotype implicatures. They have to be extrapolated, if you will, or interpolated. They have to be inserted by the hearer on the basis of Gricean implicatures.

And why is this? Because ultimately this description conforms to principles of narrative discourse, and it seems that in narrative discourse we basically only talk about the events that happen. We do not talk about why it is that

the events happen. That information is left underspecified. That seems to be a principle that is valid across languages.

Whatever the reason, this is what we got in language after language, regardless of whatever other typological properties the language had. So clearly, this approach of just showing people these clips and asking them to tell us what happened wasn't sufficient to get more data on how causal chains are encoded across languages, because all that got us were these representations of causal relations in terms of implicatures.

We therefore devised this additional study, which added to the stimuli those elicitation prompts that I hope you remember. Along the lines of 'Why did the triangle break?'; 'Was there some entity that broke the triangle?'; if so, 'Which of the entities broke the triangle?'; and so on and so forth. Now we also edited some more stimulus clips, and together these instantiated a sort of make-shift etic grid that distinguished between four different maximally possible participants in a causal chain, namely an initial causer; a causee which is an intermediate causer, or an intermediate animate link, if you will, in a causal chain; an instrument which is an intermediate inanimate link in the causal chain; and an affectee which is a final participant in the causal chain. And notice that we're using the term 'affectee' somewhat differently from how some other people in the literature on causality are using that.

So that gives us four different possible types of causal chains. The simplest type is one that involves the direct action of a causer on an affectee. The second simplest one involves a causer acting on an affectee with an instrument. The third possibility: the causer causes a causee to act on the affectee. And the fourth possibility is causer causing a causee to act on an affectee with an instrument.

In addition, we also varied an orthogonal dimension, which we called 'contact,' which has to do with whether the participants and the subevents of the causal chain are spatially and temporally contiguous. In this clip, the circle hits the square, and the square starts moving as a result of that collision, and there is no gap in space nor any lapse in time between these subevents, the hitting and the motion of the square. In this case, you have causation across a gap in space. So the circle doesn't actually hit the blue square, but the blue square does move off immediately as soon as the circle reaches the point closest to it.

Next up, the square only starts moving after a moment. There's a latency, a lapse of time between cause and effect. In another clip, you get a combination of the spatial gap and the lapse of time. This is inspired by classical work on event perception that was done by Albert Michotte and his students all the way back in the 1930s and 1940s. Michotte had the idea that causality is attributed, not at the level of abstract reasoning, as the Scottish philosopher

David Hume had claimed famously all the way back in the 18th century, but rather immediately at the level of visual processing. So there are visual stimuli that cause people to interpret them to the effect of causality because of certain visual appearances. And there are other stimuli that people tend to interpret as not involving causal relations. Basically, Michotte's theory is an attempt to apply principles of Gestalt psychology to the processing of causal information.

The final dimension that we tried to build into the etic grid is that of Force Dynamics, again inspired by Len Talmy's work. In particular, we tried to distinguish between causation and enabling type scenarios. So let's consider a dropping scenario, where you could say that *I caused the pen to fall*, but the way we conceptualize this, the primary cause is actually the force of gravity, and what I did is that I removed my support of the pen that previously was neutralizing the pull of gravity and thereby allowed gravity to take its course, to exert its full force on the pen.

And our idea was that these three independent dimensions—the complexity of the causal chain in terms of the number and kinds of participants involved, the dimension of contact, and the dimension of force dynamics—would together, but independently of one another, determine directness of causation, in a sense that the most direct types of causal chains would be represented with using the simplest causative constructions, and the least direct or most indirect causal chains would trigger the use of the most complex causative constructions or the most complex descriptions. And that's basically what we found.

This is a pilot study. Originally we did it just in preparation for looking at a much larger sample of languages, but we haven't gotten around to doing that. This is a study involving data from four languages, namely Lao, Ewe, Yucatec, and Japanese. In all four languages, the simplest type of description is the one that you get for scenes that involve only causation dynamics, no enabling dynamics in other words, and no gaps, and no lapses of time, and no involvement of an instrument. That's basically the example of the blue square hitting the red circle and the red circle going off at the very same moment.

In Ewe, Japanese, and Yucatec, you can describe this very simple type of causal chain using a single verb. In Lao, even for this very simple causal chain, you need a serial verb construction—basically something like the Lao equivalent of what would be a resultative construction in English. So you would say 'I smashed the glass broke' or 'I smashed the glass break' and it would mean that I smashed the glass and caused it to break or something like that. And this type of description of course has the Macro-Event Property unsurprisingly. So you cannot insert time adverbials that only have scope over either the smashing event or the breaking event.

So then one degree of complexity removed, you get a lapse of contact somewhere in the causal chain and a delay between cause and effect. So those are the ones where the dimension of contact is responsible for giving you a more complex and less direct causal chain. So this one can still be described with a simple transitive verb clause in Ewe and Japanese.

But the Yucatec speakers require a periphrastic causative construction, so something like, instead of 'The blue square pushed the red circle off the stage,' Yucatec speakers would say, in the case where you have the gap and the delay of time both combined, the Yucatec speaker would say something like, the Yucatec speaker would say something along the lines of 'The blue square caused the circle to move,' but not 'The blue square pushed the circle.'

The Lao speakers also would prefer, although don't strictly require, a different serial verb construction that involves something like a causative light verb. So they would say something like 'He made the glass break' rather than 'He hit the glass break,' if there is a delay of time involved here between the hitting and the breaking. But again these are all macro-event expressions, both in Yucatec and in Lao.

Then, once enabling dynamics enters the stage we use a clip of a woman who drops the hammer onto the plate, and the hammer causes the plate to break. So that would be an example that involves enabling dynamics. Actually it's not the most simple example, because that event obviously also involved an instrument, so three participants rather than just two. When enabling dynamics is involved and no contact at the time of change, this kind of causal chain can be described with a simple transitive verb clause in Ewe, Japanese, and Yucatec. So you can still say something like 'The women broke the plate.' But in Lao you need one of those periphrastic causative serial verb constructions.

As soon as mediation by instrument is involved, you need more complex constructions in Ewe, Yucatec, and Japanese. So in Ewe you can in fact still use a simple transitive verb construction, but there is a preference for a serial verb construction that works semantically like a resultative construction. In Yucatec, a causative light verb construction is now strongly preferred over a simple transitive verb description. And in Japanese, people will prefer a converb construction. 'The woman broke the dish five minutes later after dropping the hammer' is impossible in this construction. You have to go to a more complex converb construction if you want to time the time between the dropping event and the breaking event, and therefore you can say that this one still has the Macro-Event Property.

Finally we get to the involvement of a cause. The example would be one person hitting another person, bumping into another person and causing the

second person to drop something that then breaks. So this involves both a causee and enabling dynamics and lack of contact.

In Ewe, at this point, a periphrastic causative construction is at least as good as a simple transitive verb. In Yucatec, a light verb construction is preferred over a simple transitive verb clause. In Lao, the periphrastic light verb construction is the only choice. And in Japanese, a converb construction is required in descriptions of all those scenes that involve a causee. But this time it's a type of converb construction that lacks the Macro-Event Property. So this one actually allows the insertion of a time-positional adverbial with scope over a subevent only or over the distance in time between the subevents.

In Japanese, half of the scenes that we included in this study could not actually be described with a single macro-event description, a description that has the Macro-Event Property. Whereas in the other three languages, there was not a single clip, a single type of scene that triggered a non-macro-event description as the preferred response. So there is a striking contrast between Japanese and the other three languages.

Again, in Japanese, for half of the scenes, including all the ones that involve a causee, you have to use a converb construction which does not have the Macro-Event Property. So, half of the scenes cannot be represented as single events in Japanese, whereas in all of the other three languages all the scenarios in the clips can be described using expressions that have the Macro-Event Property.

So why is that? Why is Japanese so different? Basically because there are no periphrastic causative constructions in Japanese, either because the appropriate light verbs, of the type 'do' or 'make' or whatever languages use to express this concept, are lacking in Japanese, or perhaps more appropriately speaking, because in Japanese, those light verbs actually are incorporated as bound morphology. Derived causatives in Japanese, as well as in other languages, actually pattern with simple causative verbs, simple transitive verbs in terms of their compatibility with more complex causal chains.

So basically, for example, Yucatec has morphologically derived causatives. Those are used only for the simplest types of causal chains, just like simple root-transitive causative verbs. In other words, what morphologically derived causatives do in languages such as Yucatec, and apparently also Japanese, is they fill lexical gaps among the descriptors of simple causal chains, rather than to produce descriptors of more complex causal chains.

So the picture overall is similar to the one that emerged from the first case study, the one on motion events, in the sense that there is a surprising amount of variation across the languages of the world in terms of how much information can be represented in a single macro-event description, and that

information is driven by differences in lexicalization and in the availability of constructions.

Now very briefly, let's consider the third case study. I'm only going to give you a very general idea of what we've been looking at. So again the question is across languages, is there something like a macro-event phrase, a type of construction that universally is associated with the Macro-Event Property? And our point is that verb phrases and clauses, of the type that are recognized in most syntactic theories, are not a good candidate for the Macro-Event Property.

We illustrate this with an example from English: *Floyd complained from his departure in Nijmegen to his arrival in Düsseldorf*. This is a single verb phrase, and nevertheless it lacks the Macro-Event Property, because these event nominalizations *departure* and *arrival*, are still compatible with their own time positional modifiers. So you can say *Floyd complained from his departure in Nijmegen at 11:00 to his arrival in Düsseldorf at 12:30*. But in Role and Reference Grammar, these event nominalizations introduce their own cores in the syntactic representation, and these cores come with their own periphery, and the periphery accommodates those separate time-positional modifiers.

So our hypothesis is that the verbal core in Role and Reference Grammar, unlike the verb phrase in traditional phrase structure grammar, including in mainstream Generative Grammar, is an 'event phrase.' And that's not surprising if you know anything about Role and Reference Grammar. Because the definition of a verbal core is basically a lexical event descriptor, which is what's called a 'nucleus' in Role and Reference Grammar—combined with the expressions of its semantic arguments. Then a core has a periphery, which contains the modifiers of the event description, including time-positional modifiers.

So the hypothesis that we are exploring in this paper—which we've been working on for a number of years and [which] we first presented in 2009 at the RRG conference at Berkeley—is we're looking at data from a bunch of languages, serial verb constructions from Ewe, converb constructions from Japanese, event nominalizations from English, and also light verb constructions in Yucatec, in order to support a two-part hypothesis which I'm calling here the Core-Macro-Event Property Hypothesis: Across languages, single-core constructions necessarily have the Macro-Event Property, whereas multi-core constructions generally lack the Macro-Event Property, unless the nexus that combines them is one of co-subordination. So core co-subordinations across languages again tend to have the Macro-Event Property.

Now what's co-subordination? Very briefly, in Role and Reference Grammar, three so-called nexus types are distinguished: namely the traditional 'coordination', although that means not quite the traditional thing, and 'subordination', but also 'co-subordination'. Now coordination is a kind of nexus where

two expressions form a superordinate complex expression without one being an argument of the other, which is what constitutes subordination, and without the two sharing any part, which is what constitutes co-subordination. So in particular, we think there are two flavors of co-subordination in the languages of the world. Co-subordination can be constituted by argument sharing, that's the tighter form of co-subordination, or merely by periphery sharing, so two cores that together share a single periphery, that's the looser form. But both of these, for obvious reasons, cannot accommodate multiple time-positional modifiers and therefore have the Macro-Event Property.

Event segmentation can't be captured in terms of lexicalization alone, unlike the domains of the classic studies of the cognitive anthropologists, since events aren't encoded by single lexical items. That's why we have to look for new tools, if you will, new kinds of diagnostics, to capture properties of semantic typology at the syntax-semantics interface once we move beyond the lexical level.

Semantic typology of event segmentation requires a criterion that's sensitive to differences in syntactic packaging, but that's applicable crosslinguistically regardless of construction type. And the Macro-Event Property fits this bill. It assesses the event construal of a construction in terms of the temporal operators the construction is compatible with. And this construction has the Macro-Event Property if temporal operators necessarily have scope over all the subevents.

I've presented two crosslinguistic studies, one that looked at motion events, and one that looked at causal chains. Both of these have shown that lexicalization and clause-hood aren't always sufficient predictors of event segmentation, and that event segmentation varies with both lexicalization and the availability of certain event-denoting constructions.

Remember from the first study, Ewe and Lao have those serial verb constructions that allow them to string what otherwise would be verb-framed verb phrases or verbal cores together to form complex single macro-event expressions, whereas languages like Yucatec and Zapotec and Saliba and so on and so forth lack those serial verb constructions, and that's what causes them to be Type-III languages, meaning languages that require a separate macro-event expression for each subevent. And similarly Japanese, because it lacks those periphrastic causative constructions, requires a separate macro-event expression for cause and effect in any more complex causal chains, any causal chains that involve a mediating causee in particular. There is a little more, there's a somewhat more complex story to tell, which I didn't get around to explaining. So there are conditions under which you can do that in a single simple macro-event expression in Japanese, but then you're attributing intentionality to the

initial causer. So you are saying that the initial causer intended to cause the ultimate outcome, which often isn't the case in the clips that I showed you.

These studies have also confirmed that the extent of crosslinguistic variation in event segmentation is indeed that which Andrew Pawley originally pointed out just on the basis of his work on Kalam, which, as I said, formed historically a source of inspiration for this work.

We discovered some constraints on form-to-meaning mapping, such as the unique assignment of thematic relations or semantic roles in macro-event expressions, which seem to be sensitive to the Macro-Event Property and thereby suggest that this Macro-Event Property is a property of the design of language itself and not merely an ad hoc criterion that semantic typologists can use as a diagnostic.

I suggested that simple cores in Role and Reference Grammar are macro-event phrases. They appear to be universally associated with the Macro-Event Property, unlike verb phrases. And cores, but not their nuclei, license the kind of periphery that accommodates time-positional modifiers, which explains this correlation, and cores are the smallest syntactic unit that can host a syntactically complete event description. Or at least that's our hypothesis.

LECTURE 10

The Language-Specificity of Conceptual Structure: Taking Stock

So what exactly do we have in store for the last lecture? I'm going to try to do two things. First of all, I'm trying to wrap up, in a sense, which is a strange thing to do. That's an anti-iconic order. I'll start from the end, by giving you my sense of how much variation there is across languages in semantic categorization. I do not really know the answer to the question how much variation there is, and I'm actually very glad that I don't know the answer to that question, because that means I can go on working on the stuff I am really enjoying working on for a while. Then I'm going to present one more study, one more line of research that I've been pursuing for a good ten years, and that leads a roundabout way from the representation of motion in language to questions about language evolution and the relationship between language and cognition, nonlinguistic conceptualizations.

So first of all, here is my take on what's variable and what's uniform across languages in how they represent the world, or reality. I'm going to take as my point of departure the article by Evans and Levinson (2009) in *Brain and Behavioral Sciences* that I already mentioned several times before during these lectures. So to recap, the case that Evans and Levinson are trying to make is they are trying to educate the cognitive sciences outside linguistics, and to some extent within linguistics, about the extent to which linguists have been able to find language universals.

The point is simply that there is a widespread assumption in some areas of linguistics and especially outside in fields that are related to the language sciences, such as computer science, artificial intelligence, psychology, and so on, that actually languages are, all in all, really similar to one another. And this is a view that has been propagated by linguists starting from Humboldt all the way to Chomsky. So, Humboldt famously said that one possible view of the

All original audio-recordings and other supplementary material, such as any hand-outs and powerpoint presentations for the lecture series, have been made available online and are referenced via unique DOI numbers on the website www.figshare.com. They may be accessed via this QR code and the following dynamic link: https://doi.org/10.6084/m9.figshare.11419152

© JÜRGEN BOHNEMEYER, 2021 | DOI:10.1163/9789004362628_011

languages spoken on the planet is that you could think of them all as dialects of a single language. And the approach that Noam Chomsky developed—the framework of Universal Grammar—very much pursues that same perspective.

As opposed to this, Evans and Levinson make a strong case that, as a matter of fact, the empirical evidence suggests that there are possibly no language universals in the sense of absolute, exceptionless universals along the line of 'All languages have X.' Whenever you start out with a generalization like that, whether it's over the speech sounds, whether it's about the syntactic structure, or the semantics of the languages of the world, you're going to find exceptions. So variation is much larger than is widely assumed in the cognitive sciences.

And Evans and Levinson advocate a view of language not as a strongly biologically determined system but rather as what they called a 'bio-cultural hybrid,' meaning something that has roots in biology that are encoded in the human DNA, but the development of which is informed to a very large extent by culturally transferred knowledge. That's something I strongly agree with.

What I *dis*agree with is the emphasis on variation, on the absence of strong universals, which is correct while at the same time, however, neglecting to mention that there is an amazing amount of recurrence of patterns across the languages of the world in syntactic structure as well as in semantics, and that's something that calls for explanations.

To give you an example of what I have in mind here, take tense, something that I worked on for my dissertation and that I'm still very much interested in and continue working on. So there are languages that have tenses. A lot of languages have tenses, and there are languages that have no tenses. There are languages that have viewpoint aspect systems. And there are languages, such as my own native language German, that at least in the standard variety don't grammaticalize aspect. There are languages that combine tenses and aspects in a single functional category system. There are languages that keep them apart, and so on and so forth. So there is a lot of variation.

Nevertheless, the number of tense-aspect systems is quite limited across the languages of the world. There are certain architectural trends in tense-aspect systems that recur over and over and over again, similar categories that you find in language after language after language, in unrelated languages spoken in different parts of the world, and so on and so forth. That's something that needs explaining.

Moreover, and more intriguingly for me, the semantic principles underlying the interpretation of utterances in tensed languages, meaning languages that do have grammatical tenses, are at the core identical to the principles that govern the temporal interpretation of utterances in tenseless languages such as Yucatec Maya, the language that I've been working on for twenty years, or,

for example, Mandarin. If you look at the inferences that people derive in discourse, when they are trying to understand the temporal references that the speaker is trying to make, they follow the same principles. So there is an underlying semantic system here that is expressed in some languages, and not in others. And that's something that calls for an explanation.

Now the standard line of reasoning about this kind of phenomenon in the cognitive sciences used to be that there is an innate conceptual category here that is shared among all human populations, whether it's expressed or not. When it is linguistically expressed—meaning in those languages that do have expressions lexicalized or grammaticalized expressions—then the underlying conceptual categories inform the semantics of the lexical or functional categories that exist in the code of these languages. But if the particular language has no expression, the conceptual categories are still the same in the minds of the speakers.

Over the last 20 years there has been an overwhelming amount of evidence suggesting that this view is too simple. First of all, it's not the case that the underlying conceptualizations are independent of the languages spoken by the speakers. Rather, there is now quite a lot of evidence suggesting that the structure of the language influences the underlying conceptual representations. And secondly, I think we should stop taking it for granted that categories' conceptual distinctions that are recurrent across languages necessarily have to be innate. We should consider other possible explanations.

Now, I've talked about one possible explanation that was suggested in an article that appeared in the journal *Nature* last year, authored by Michael Dunn and colleagues, in which they don't talk about semantic typology, but syntactic typology. But the argument they are pursuing is that the patterns of word order typology that have been keeping syntactic typologists busy since the days of Greenberg and for which my colleague Matthew Dryer found statistical evidence based on a very large database—that these patterns in fact primarily tend to agree among languages that are genealogically related, meaning that have a common ancestor, and tend to vary much more across language families.

From this, Dunn and colleagues developed the argument that, well, maybe typology really isn't able to teach us anything about the design of possible human languages, the structure of the design space for human language, and therefore the place of language in cognition, because all typology really can do is describe what has come down to us from the common ancestor of all human languages that are still spoken on the planet.

I have several problems with this argument. I do very much appreciate this new line of research. I definitely think that this is a part of the picture that we

need to investigate. However, I'm actually a little bit skeptical about monogenesis, meaning I don't think we should take it for granted that all currently spoken modern languages have a common ancestor. As a matter of fact, we know that that's not the case, because there are lots of languages that have originated more recently as contact varieties—pidgins and creoles.

Equally important, I think the argument whether monogenesis is a viable explanation for most of the languages of the world or not is oversimplified. There are tendencies across languages and properties that do allow us to directly infer properties of the underlying cognitive system and therefore make inferences about the design space for human language. The study I'm going to try to present at the end of today's lecture on the representation of motion paths in Yucatec actually is an illustration of this.

Now let's take a quick look at the principle domains—the principles of phenomena of the syntax-semantics interface—as the main area of study for semantic typology, and take a look at what we know now about what's variable and what isn't across languages. I'm going to break this down across four kinds of phenomena, four domains, if you will: lexicalization; lexical categories or parts of speech; functional categories, meaning function words and inflections; and semantic composition, the principles underlying the computation of sentence meanings.

First, lexicalization. I mentioned briefly the research that's been conducted in recent years by Terry Regier and colleagues at Berkeley—originally Chicago, now Berkeley—on what you might want to call the naturalness of the concepts that get lexicalized across languages. In other words, the semantic categories that are expressed by lexical items, in particular in the domain of color terminology and, most recently, in the article that appeared just in the last two weeks in *Science*, the article by Kemp and Regier in the domain of kinship terminologies.

What they found in most of these domains is that the categories expressed in language after language tend to be natural in a sense that they maximize similarity within categories and minimize similarity across categories. They tend to aim for an optimal balance between communicative efficiency, where the goal is to make it easy for the interlocutor to understand what category you are putting a referent in, and cognitive efficiency, where the goal is to minimize the amount of effort that it's going to take the speaker to pick the right category. I am pretty certain that this will hold true for other domains of lexicalization as well. By the way, by lexicalization I mean the availability of words for concepts—lexical items that express concepts across languages.

I also think that there are other similar principles that people will discover and have been assuming sporadically in the past. For example, natural

categories tend to be contiguous with respect to the way they partition an underlying domain. If you go with that Munsell color chart, which goes by brightness and hue, to the extent that we can assume that as a valid representation of the underlying conceptual space—and it's now widely agreed that we can't, that this is not a good model; but that aside—you are not going to find a language that uses a color term for very different from each other, while colors close to each other in those values would be considered the same.

Of course, there are similar lines of thinking that go back to the generative semantics debate. For example, there is a long-standing argument in logic that certain kinds of quantifiers get lexicalized across languages and others don't. People have long been talking about possible verbs, for example, in terms of the causal chains they lexicalize, and the underlying principle is always the same. You can always boil it down to something like naturalness. There are categories that are just too goofy—too funky to get lexicalized across languages. I'm going to leave you with a general constraint against the lexicalization of funky stuff.

This obviously does not mean that any of these so-called natural categories that get lexicalized are in fact innate. Quite on the contrary, what this suggests is what I was already hinting at on Monday, which is that which is shared cognitively by speakers of different languages and which is responsible for this design of natural semantic categories are the principles of categorization, not the content of the categories. However, it is unquestionably also true that there are limits to the kinds of concepts that get lexicalized, possibly beyond even this complexity constraint that I was alluding to as funkiness.

Again, my current thinking is that it may very well be possible to try to explain these kinds of patterns as well in terms of an interaction between a general innate categorization engine and environmental data, where these environmental data include data that are transmitted by the members of the culture, be it through the use of language or be it through other means. During the lecture on spatial frames of reference, we talked a little bit about how that might work.

Dimensions of variation that we encounter in lexicalization concern the domains that get lexicalized. That's the content I mentioned. For example, like I said, not all languages have basic color terms. Not all languages, in other words, lexicalize color at the basic mono-morphemic level. Not all languages have CUT-type verbs. That's something that emerged from this study of verbs of cutting and breaking that all languages seem to have BREAK-type verbs, meaning basic state change verbs that can be used to describe this domain of separation in material integrity, to use Ken Hale's term. Not all languages have CUT-type verbs, meaning that not all languages have instrument-specific verbs.

And this has been related to the fact that not all cultures—this is an argument courtesy of Steve Levinson—not all cultures until recently used to have a lot of manufactured tools, even relatively primitive tools. They may have primitive tools, but not steel tools, and so on and so forth. These are pre-industrial cultures—and pre-industrial culture is something that is now vanishing from the planet rapidly. Until a little more than a century ago, there used to be many pockets of pre-modern culture, including material culture, and that seems to be responsible for variation in what gets lexicalized.

There's variation in how extensively and intensively domains are lexicalized, where by extensively I mean the total number of terms. So if you think of the domain of spatial dispositions, European languages typically have 'sit,' 'stand,' and 'lie,' a few posture verbs, and that's pretty much it. Then you find sporadically other concepts lexicalized in various classes of verbs, but that's it. Whereas Mayan languages have these hundreds and hundreds of roots that make up a form class, and that talk about spatial configurations such that in Yucatec, for example, you can lexically distinguish between, let's say, seven different kinds of hanging, seven different kinds of leaning, and so on. There are intensive differences in terms of network density of the relations, the lexical semantic relations among these items.

Moving on to lexical category systems, or parts of speech, so the argument that has—I don't know who first made this argument formally, but it's something that has long impressed me. It's an argument that has been made both for lexical category systems and for functional category systems, meaning function words and inflections—it's very similar to the no-funky-stuff constraint. You don't find languages that have a separate lexical category for temperature. You don't find languages that have a special lexical syntactic category that expresses nothing but concepts of temperature. So far, such a language has not been attested, and I'm prepared to offer you a bet that you won't find such a language. There are many other concepts that will not surface in dedicated lexical category systems.

At the same time, there is a ton of variation in the kinds of lexical categories that languages have. One example of that is obviously the category of numeral classifiers, which almost don't occur in Indo-European languages, but they do occur in Mandarin, in Yucatec and in many other languages. The domain that I mentioned a moment ago, of roots that describe spatial dispositions in Mayan languages or Iroquoian languages and elsewhere in the world, is another example.

Many languages of Australia have a special lexical category that combines with verbs and expresses the bulk of the meanings that are expressed by verbs in better-studied languages. But those same languages have only very few verbs.

This is something that's been found in many languages of northern Australia, also in languages of New Guinea, and more recently in languages of the Andes. These co-verbs are a lexical category that doesn't occur in other languages. So, yes, there definitely is variation, and we still discover new lexical categories that grammarians previously missed.

At the same time, it's clearly *compact*. It's a compact variation space. If I say *confined*, I'm suggesting that there are strong constraints on the kinds of lexical categories that a language can have.

So let me restrict myself to saying that the variation is a lot less than it could be, and of course here Dunn and colleagues would argue that that's just simply a reflection of the fact that language evolution, as it has proceeded over the past at most one hundred thousand years that most people now think that modern languages have been around, simply hasn't been able to explore more than a relatively small portion of the design space for possible human languages, just simply because that area that's been explored is going to grow as a function of the amount of time that language evolution has had, but that function is very slow. So you have to add maybe a million years, and then you get a much better sense of what the possible design space would be like, except by then probably the earth has crashed into the sun or something like that.

I'm going to look very quickly at a paper that Matthew Dryer presented at a variety of talks some ten years ago but that he never published, in which he claimed that there are limits to, let me say, where languages draw the dividing line between nouns and verbs and adjectives. So he started out from a notional space of nine conceptual classes, where all the dynamic concepts—'throw' and 'run' and 'die'—go into verbs, they are lexicalized as verbs. And these ones, all the stative concepts, 'boy,' 'father,' 'love,' 'dog,' 'large,' and 'house' are lexicalized as nouns and/or adjectives. So he said there's no language that actually does this.

Similarly this one is a cut that separates all the dyadic from the monadic concepts, so 'boy,' 'dog,' 'house,' 'die,' 'run,' and 'large' are all monadic, they have a single semantic argument, whereas 'father,' 'love,' and 'throw' are dyadic or transitive, whatever you want to call it, so they have two semantic arguments. Dryer said there is no language that separates verbs from other parts of speech purely on the basis of this division.

So he presented this talk in Amsterdam when I was still at the Max Planck Institute in Nijmegen, and I went and attended his talk, and I pointed out to him afterwards that actually Yucatec does exactly this. 'Die,' 'run,' and 'throw' are verbs in Yucatec, whereas 'love' is a relational noun in that language. And 'large' is an adjective, or a 'stative predicate,' as I like to call it, but stative predicates and nouns share a number of morpho-syntactic properties that separate

them from verbs. So there's a deep split in the architecture of the lexical categories of this language that works exactly this way.

There is in fact variation in the semantic principles underlying the distinction of parts of speech, very significant differences across languages in terms of the underlying principles that divide lexical categories from one another. And so, given this variation, the question of why exactly languages have the lexical categories they do is ultimately a question that I consider unanswered and that I think is one of the more important questions that semantic typologists and linguists in general ought to try to answer, because obviously the architecture of the system of lexical categories is very important for the grammar as a whole.

Let's finally consider functional categories. Well, this is in a sense what I started out with. I talked you through what I have to say about tense and aspect—I did say already that the no-funky-stuff constraint applies to functional categories as well. So there is no language, as far as I know, that has an inflectional category for, let's say, temperature. So no language has grammaticalized temperature as an inflection in verbs or nouns or whatever. No language has grammaticalized flavor as an inflection. No language has grammaticalized color as an inflection, and so on and so forth. Obviously there's got to be a reason for this, but we don't quite yet understand the reason for it.

A lot of people have pretty good intuitions. I think it's a pretty good guess that functional categories are what they are because they play a sort of supporting role in language. So the bulk of the content that we try to express using language is expressed in lexical categories, and the functional categories somehow serve to disambiguate what we mean by a particular array of lexical items.

How exactly that works is something that you are only going to be able to answer through collaboration between psycholinguists trying to uncover the role of functional categories in language production and comprehension, typologists trying to uncover the amount of crosslinguistic variation in functional category systems, historical linguists trying to uncover the principles that cause functional categories to grammaticalize and change, and language acquisition researchers who study how kids learn functional categories and the assumptions that they initially make about the functional category systems of their native languages. So only through a huge joint effort of all these different subfields are we going to be able to come up with an answer to this question.

The problem with the no-funky-stuff constraint is define "funky stuff." There's an amazing amount of variation, for example, in determiner systems across the languages of the world. Some languages conflate evidential distinctions—distinctions of visibility, audibility, and so on—in the determiner system. Some languages conflate information about spatial proximity

in the determiner system. Some languages conflate posture in the determiner system. That used to be something very exotic. One of the early reports was by Harriet Klein from the languages of the Andes. By now people have found examples from languages all over the world that happen to combine posture roots with demonstratives to grammaticalize determiners, so you get different definite articles for referents of different shape or referents of different axial structures and so on and so forth. Why is that? Why does that seem an obvious thing to do, and even if the number of languages that do it is very small, if languages all over the world do it, it must still be in some sense an obvious thing to do. Why is that? It calls for an explanation.

Semantic composition is the topic I'm going to have the least amount to say about. But my main message is there has been a lot of really exciting research in this area over the last ten years. So until about ten years ago, it was assumed that Frege and Montague had figured out how semantic composition works. Do we assume that semantic composition is something that really happens in natural language, or do we assume that this is something that is an idealization. Nobody seriously questioned the universality of the principles.

Now that's changing. Sandy Chung and Bill Ladusaw suggested in a book they published in 2004 that different languages actually have different rules that they use for combining simple meanings to form larger meanings. My colleagues Jean-Pierre Koenig and Karin Michelson have been working on argument structure in Oneida, an Iroquoian language, and they argue that this Iroquoian language—or maybe Iroquoian languages in general, maybe even one subtype of polysynthetic languages in general—has a completely different way of doing argument satisfaction. They separate systematically the function of argument satisfaction and the introduction of new discourse referents. Argument satisfaction is something that's expressed morphologically in the verb, whereas the introduction of new discourse referents is done by nominal, which are systematically not syntactic arguments.

There is an unpublished paper by me and some former students on the semantic composition of motion descriptions across languages of Mesoamerica, where there is quite a bit of variation, which is largely driven by variation in terms of where the particular language expresses path information, whether the language has adpositions in the English sense at all, and so on.

The absence of absolute universals strongly suggests that semantic categorization is not innate. At the same time there's evidence that semantic categorization is subject to cognitive design principles that may well partly be innate, or perhaps more appropriately are thought of as being the product of an innate categorization engine that produces generalizations over an input that's determined both culturally and by the physical environment and by

individual experiences. So in other words, this is exactly what we should find if we take the idea of language as a bio-cultural hybrid seriously, something that is shaped by both the forces of natural evolution and the forces of cultural evolution. And that holds not just for language of course, but also for the human mind in general.

I want to take a quick look at one more study, which is going to take us back one more time to that Mayan language that I've been working on and to which I owe my career. Not just the language, but the speakers in particular, I might say. And I want to look one more time at the domain that has occupied most of these lectures, and that is the domain of the representation of space.

The question is quite simply how much spatial information gets represented in language? Which is a question that has kept a lot of researchers busy, not just since that classic book by Miller and Johnson-Laird, *Language and perception* from 1976—in fact it has been occupying philosophers and scientists for centuries. This time I'm going to look at motion events. I'm going to be interested in particular in information over the path of the motion event, where the path originates in a source location and terminates in some goal location and in between there is a route, and all these things are typically described in language in terms of reference entities that define them, or referential grounds.

I'm going to make three assumptions by way of background for the following. I'm going to make the assumption that there are at least two different systems of internal cognition that encode spatial information. One is a system that is symbolic, has algebraic structures, a system of symbolic representations similar to those of language. So you may think of the system in which linguistic semantics is actually encoded, which is what Ray Jackendoff has been proposing. Now Jackendoff likes to draw these neat boxes across faculties of cognition, which is supposed to represent modularity. I'm actually skeptical about this idea. I don't think that the mind is very neatly modular, but I have to admit this has the advantage of generating a lot of testable and falsifiable claims. That's the advantage of strong hypotheses.

Then there is a Conceptual Structure system, which allows for symbolic representations of spatial relations; but there is also what Jackendoff calls a 'Spatial Structure' system, which supports image-schematic representations, representations similar to Johnson-Laird's mental models, for example, or the image schemata that Langacker has been talking about. The Spatial Structure system is, according to Jackendoff, more directly tied to the visual and haptic representation system. So this is for me fundamentally an iconic system of representations, whereas others are symbolic representations.

My second assumption is that the representation of spatial information in Spatial Structure, the iconic system, is much richer than in the Conceptual Structure system. And this is an idea that philosophers have been pursuing

all the way since the days of Bishop Berkeley. Clearly iconic representations of space are more powerful than symbolic representations of space. It is very difficult actually to break down space into portions that can be symbolically represented. Space is not a domain that easily submits to symbolic encoding. Jackendoff's famous Conceptual Structure formula for path functions is as follows: FROM AT tree, VIA AT lake, TO AT hill, so that happens in the Conceptual Structure system.

My third assumption is that any kind of spatial information that's encoded at Conceptual Structure must also be interpreted at the Spatial Structure system, and that the opposite does not hold. Why is that? Because spatial information must be interpretable visually, so the Spatial Structure system has to be involved. But clearly there is a lot of nonlinguistic processing that goes from visual representations to inferences derived at Spatial Structure and then back to haptic output that never involves language, and therefore there's no particular reason why it needs to involve Conceptual Structure. This is an important point for the argument that I'm going to make.

The question is how much information is encoded at Spatial Structure only, and how much information is duplicated in the Conceptual Structure system. Jackendoff argued that path functions have to be conceptually encoded in Conceptual Structure, and in fact he made the encoding of path functions in Conceptual Structure part of the very core of Conceptual Structure, in a way that we'll see in a moment. I'm going to try to make a case that path is not universally encoded in Conceptual Structure, but that there are apparently populations that encode path at Spatial Structure only, in other words, in the iconic image-schematic system only. In fact, the encoding of this alleged supposed putative core function of Conceptual Structure is in fact language-specific.

That in turn is going to support an argument that rather than language having evolved as an external expression of Conceptual Structure, as Jackendoff claims—Jackendoff has been arguing for a while that Conceptual Structure evolved a long time ago. It's shared between humans and all higher animals, and language as we know it is a relatively recent development that is basically an outgrowth of Conceptual Structure. Therefore language has the structure that it does because that structure more or less reflects the structure of Conceptual Structure.

I think this is a possibility, but I don't think it's true. An alternative possibility that I want to put on the map—within this model of Jackendoff's, if we, for the sake of argument, assume this model—is that, in fact, what we share with all higher animals is the Spatial Structure system, the image schematic system. Conceptual Structure actually is something that evolved only very recently, and it evolved as an interface between language and internal cognition. In other words, these symbolic concepts basically emerged as a solution to

the problem of how to translate these Spatial Structures, these iconic Spatial Structures, into the kinds of symbolic representations that we need in order to communicate among one another.

This is why, according to Jackendoff, the encoding of path functions forms part of the core of Conceptual Structure. This is due to the Thematic Relations Hypothesis, which says that actually, most Conceptual Structure functions in abstract semantic fields are metaphoric uses of spatial representations. So, this is the more general claim that is familiar to us under such labels as 'embodied cognition' in the Cognitive Linguistics literature. Jackendoff is recasting the same hypothesis in terms of this Thematic Relations Hypothesis, which says that fundamentally space is the model for the cognitive representation of non-spatial domains. So more concretely:

> In any semantic field of [EVENTS] and [STATES], the principal event-, state-, path-, and place-functions are a subset of those used for the analysis of spatial location and motion. Fields differ in only three possible ways: a. what sorts of entities may appear as theme; b. what sorts of entities may appear as reference objects; c. and what kind of relations assume the role played by location in the field of spatial expressions.
>
> JACKENDOFF 1983: 188[1]

Now it is not the case that Jackendoff, as far as I know, has ever made explicit claims about what's innate about Conceptual Structure and what isn't. But it's fairly clear that his view submits to an interpretation under which the overwhelming number of the everyday Conceptual Structure representations that people entertain are in fact learned and culture-specific. There are only very few primitives that are innate and that are part of the architecture of the system itself.

In actual fact, this is what most people think. The most prominent exception is Jerry Fodor, who famously believes that all concepts are innate, and that is a view that most people find very hard to fathom. That seems to include Jackendoff. So the bulk of conceptual representations would be learned and culture-specific, but there is a core system that would have biologically evolved and would be coded innately. And since Jackendoff is arguing that metaphorical spatial relations, and that the thematic relations system is part of the core of Conceptual Structure, I'm going to make the reasonable assumption that it would then follow that this would be part of what's innate and universal about Conceptual Structure.

---

1 Jackendoff, R. 1983. *Semantics and cognition*. Cambridge, MA: MIT Press.

Jackendoff [1983] argues that path functions must be encoded at Conceptual Structure for two reasons. First, cognitive necessity: we need them in order to reason about this stuff. But that argument actually fell by the wayside when Jackendoff finally introduced the spatial representation system, the Spatial Structure system, around 1987, meaning the second, image schematic, module of cognitive representations, which of course supports reasoning and therefore we can reason about motion paths just on the basis of the Spatial Structure system alone. We don't actually need the symbolic representations to do that.

The second argument is linguistic necessity. Path functions must be encoded at Conceptual Structure because they are expressed in English. Let's say you have something like *The ball rolled to the hill*. This [pointing at formula on slide] is the way Jackendoff in his preferred notation represents the presumed underlying representation in Conceptual Structure. He is aware of an alternative possible representation in terms of state change, where you have an inchoative function and something that effectively means roughly 'The ball came to be at the hill.' Well, you are describing the fact that there was state change and that the outcome was the ball ending up at the hill. You are not actually encoding conceptually or linguistically translational motion of the ball along a trajectory that ended at the hill.

Jackendoff gave a number of arguments for why this can't be the underlying conceptual representation for this sentence, *The ball rolled to the hill*. You see the conceptual representation for *The ball split*, which is a state change representation. The question is why isn't the meaning of *The ball rolled to the hill* like the meaning of *The ball split*, which would be represented by the CS formula. He argues that, first of all, while motion with source path functions and goal path functions alone may be represented in state change terms, motion with route path functions cannot easily be represented in these terms.

So if you have something like *The ball rolled into the box*, you can represent that as 'The ball rolled and ended up inside the box.' If you have *The ball rolled out of the box*, you can represent that in terms of state change along the lines of 'The ball came to be outside the box.' If you have something like *The eagle soared across the canyon*, or *The train went through the tunnel*, *The expedition crossed the river*, *The horse jumped over the fence*—that's difficult to represent in terms of state change. Because the crucial bit here is not the source of a target state, but rather the transition between the two. I actually added a similar second argument to this one. In a sense, I played devil's advocate—or Jackendoff's advocate—and made his argument even stronger before I took it down. So composition of complex path functions as in *The ball rolled from the tree to hill*, that is also something that is not that easy to represent in terms of simple state change encoding, because now both the source state and the target state of the state change event are specified.

In addition, there is 'fictive motion' in Talmy's terms, as in *The highway extends from Denver to Indianapolis*, or *The house faces away from the mountains*. So these kinds of descriptions, on Jackendoff's account, involve uses of path functions and state descriptions. So that means there is no way of representing this in terms of state change, which in turn means there is no way of representing all path functions in terms of state change.

So the point that I want to make is, I think these arguments are valid for English. I think Jackendoff is right as far as English is concerned, but they are not valid for all languages. So I want to show that these arguments don't hold for Yucatec, and my conjecture is that Yucatec speakers don't encode path at the Conceptual Structure level, but instead rely on the Spatial Structure system in order to reason about motion. The implication of this would be that path functions aren't universals of Conceptual Structure, and therefore that what Jackendoff has argued to be a core component of Conceptual Structure appears to be in fact language-specific, which I then try to parlay into an argument for Conceptual Structure itself as being something that may have coevolved with language.

Obviously I don't want to, and shouldn't, take the time to take you through the entire argument, so let me present the key evidence. We already talked about satellite-framed versus verb-framed languages this morning, and I made the point back then that what are commonly considered core examples of verb-framed languages, such as Spanish, Turkish, and Japanese, are actually arguably better thought of as double-encoding languages, which lexicalize path both in the main verb root and also reflect it outside in for example an adposition, such as in these Spanish descriptions: 'The cart was inside the box', 'The cart entered the box'—both employ the same preposition; but for the ablative function, 'The cart exited from the box,' you have to use a different preposition. So the path is in fact reflected outside the verb root as well.

This is not the case in Yucatec. In Yucatec you use the same preposition. And there are a variety of choices that you have. You can use a preposition that means 'in' or the generic preposition *ti,'* which has no concrete spatial meaning whatsoever. Both of these are compatible with a locative description 'The cart, it is in the box,' but also with allative or goal path descriptions, 'The cart, it entered the box,' and with a source description, 'The cart, it exited, literally in the box.' So that means, if there is path encoded in Yucatec at all, then it has to happen exclusively in the verb root. It's not happening anywhere outside the verb, unlike in Spanish and Japanese and so on.

As a matter of fact, it doesn't happen in the verb root either. And I show it using those clips, some of which we've already seen. Nevertheless, in Yucatec you can say 'The ball entered the enclosure.' The description becomes better,

if you defeat the stereotype implicature that, when you say 'The ball entered the enclosure,' you mean it was the ball that moved. Stereotypically for Yucatec speakers, just as for English speakers and, I presume, for Mandarin speakers, objects change location by way of movement, but that's not necessarily the case. They can change location with respect to other objects by these other objects moving, or even by them coming into being or going out of existence, appearing and disappearing in a given spatial configuration.

As I briefly mentioned on Tuesday, I think this phenomenon was first attested by my former colleague Sotaro Kita for Japanese, but only with respect to the Japanese verbs meaning 'enter' and 'exit'—*hairu* and *deru*. In Yucatec it's a more widespread phenomenon. It holds for most of the path verbs of Yucatec. It's also possible in Yucatec to say 'The ball ascended the slide rolling' even though it was actually the slide that slid under the ball. So more systematically in Yucatec path verb roots are applicable to location changes, even if it wasn't actually the figure, meaning the theme of the motion verb, that changed location, I mean, that moved. So there is a location change in the sense of a change of the locative relation with respect to the ground, but it was actually the ground that moved.

These kinds of descriptions are possible in Yucatec and from that I conclude that the verb roots don't encode path information, either. There is a caveat that, and this is a well-known phenomenon among the people who have looked at this across languages. Some of the path verbs are much more likely to occur without figure motion than others. 'Enter' and 'exit' are the most likely ones. 'Ascend,' 'descend,' 'rise,' 'fall,' and 'pass' also do it, whereas 'come,' 'go,' 'leave,' 'arrive,' and 'return' don't do it in Yucatec. Native speakers will always tell you, no, you can't say that if no figure motion is involved. For now I will just simply say there is an alternative to the assumption that these verbs aren't applicable without figure motion because they in fact encode the figure motion.

One of Jackdendoff's arguments concerns route paths. As a matter of fact, in Yucatec the only way of expressing location change with respect to a route path is by use of the verb *máan*, 'to pass.' So whereas in a satellite-framed language like English, you have this rich set of options *across*, *through*, *over*, and whatnot, in Yucatec you have to express it using 'pass,' and I already showed you the example that illustrates that in Yucatec you can use *pass* without figure motion. The argument that the state change decomposition is not possible for route paths simply doesn't seem to be valid for Yucatec.

What about the composition of complex path functions? That was my enhancement to Jackendoff's argument. Of course, the answer to that one is actually, the composition of path functions doesn't happen in Yucatec. Because in

Yucatec, you just cannot say 'It rolled from the blue square to the green triangle.' You remember that from this morning, from the study on event segmentation. As a matter of fact, you have to say 'The ball departed from the blue triangle, and it rolled, and then it reached the green triangle.' So there is no composition of the path functions either. And here you see an actual description of that clip.

What about the argument from fictive motion metaphors? As a matter of fact, Yucatec has what you may want to call fictive location change metaphors, but not fictive motion metaphors. So you can say something like 'This road here, it exits the town of Señor, and when it has done that, it passes Tixcacal, and when it has done that, it arrives in the town of Yaxley.' So this is a Yucatec equivalent of saying *The road goes from Señor via Tixcacal to Yaxley*. You can say that. You can break it down across three clauses and describe this metaphorically, using three state change verbs. But any path function that cannot be represented as metaphorical state change just isn't encoded as a metaphorical path function in Yucatec.

Think for example of a vision path, such as, let's say, 'You looked through the window into the house,' which is a metaphorical use of the path functions 'through' and 'into.' The way you would say that in Yucatec is 'You looked at the window and then you saw what was in the house.' There is no other way of expressing that. This is an argument that is not entirely original. Yo Matsumoto made a similar point in the 1990s about Japanese, so there seems to be a larger story here. The motion metaphors that different languages use for talking about non-spatial domains actually depend on the resources that these languages use to talk about space itself. So the resources the language uses to talk about space itself determine the kinds of spatial metaphors that the language uses for non-spatial domains.

I made that same point at a more abstract level during my dissertation research, where I pointed out that Yucatec also doesn't have words such as *after* and *before* and *while*, temporal relators which often have been analyzed as a particular kind of metaphorical path relators. As a matter of fact, Yucatec talks about the order of events, or the order of time intervals, in a very different way. It talks about them basically on the basis of aspectual concepts, using verbs such as 'begin' and 'end,' and you've seen that several times. For example, for something that we would translate in English as 'and then,' the language actually uses something based on the aspectual verb 'to end': 'X happens, that ends; Y happens, that ends; Z happens'—that's the Yucatec way of saying 'X, and then Y, and then Z'. So this seems to extend to metaphorical representations of time.

There is one more source of evidence, which I have not yet mentioned. (And now we're stuck. I wonder why that is.) There is one more source of evidence,

which concerns the argument from temporal relations, and then this … [technical problems with the slide show]. If there are no path functions in the Conceptual Structure of Yucatec speakers, then that should also mean that they have a hard time learning to express the path functions of their second language, namely Spanish. And indeed in spontaneous observation, you often encounter Yucatec speakers producing errors of these kinds. They would say instead of *De donde vienes?* 'Where did you come from?' using the ablative preposition *de* that monolingual Spanish speakers would use, *Dónde vienes?*, literally 'Where do you come?' And they would say instead of *El ratón salió de su agujero*, 'The mouse exited from its hole,' something like 'The mouse exited in its hole.'

A few years ago I did an as yet unpublished study together with Rodrigo Romero, where we collected descriptions of motion events by Yucatec speakers using Yucatec, Yucatec speakers using Spanish, and English learners of Spanish, and monolingual Spanish speakers. We found that Yucatec speakers indeed produced these characteristic errors with the encoding of path functions more frequently than English learners of Spanish do, and more than monolingual Spanish speakers of course, supporting the conjecture that at the level of Conceptual Structure, path functions aren't encoded.

Of course, whenever you say a language doesn't encode an X or Y or Z, the question is, well, but surely speakers of this language must be thinking about this stuff? And of course Yucatec speakers, just as speakers of any other language, have to reason about motion paths, and have to memorize motion paths, and so on and so forth. So how do they do it? My conjecture is that they do it on the basis of the Spatial Structure system alone. The argument that I make is suppose it is indeed the case that this core part of Conceptual Structure—namely the representation of path functions, which supposedly, via metaphorical applications to non-spatial domains, is part of the basic mechanics of Conceptual Structure, the core of Conceptual Structure—suppose this system nevertheless does not occur in the Conceptual Structure of Yucatec speakers. Then to me that suggests that there isn't very much about Conceptual Structure that isn't language-specific.

If it is the case that Conceptual Structure, even at its core, is basically a language-specific phenomenon, and that would fit with an evolutionary scenario where Conceptual Structure, rather than to be the underlying model for the evolution of language, being a cognitive system that has co-evolved with the language as an interface itself that mediates between language and the image schematic cognitive representations of the Spatial Structure system. That is my argument. As far as I know, I'm not aware that Jackendoff has been very impressed with the argument.

So this is my summary for the lecture. The evidence from semantic categorization fits the view of language as a bio-cultural hybrid. The absence of absolute universals strongly suggests that semantic categorization isn't innate. At the same time, there is evidence that semantic categorization is subject to cognitive design principles that may very well be partly innate.

Motion is systematically framed as state change in Yucatec. Path functions aren't encoded. Evidence for that are path-neutral ground phrases and the compatibility with non-figure-motion scenarios. Jackendoff's arguments for the necessity of a path semantics for English don't apply to Yucatec. There are no fictive motion metaphors. Descriptions of motion with respect to route grounds are drastically underspecified. There is indirect evidence for the absence of path functions from the Conceptual Structure of Yucatec speakers that comes from the lack of temporal connectives expected to be based on path metaphors. These apparent transfer errors that L2 Spanish speakers whose native language is Yucatec make support the hypothesis that path functions aren't encoded in Yucatec Conceptual Structure, and therefore that Yucatec speakers have a harder time learning and processing path functions when they are using language.

Let's consider some implications for the architecture of cognition. The encoding of path information at the Conceptual Structure as opposed to the Spatial Structure system may be language-specific. Via the Thematic Relations Hypothesis, that would entail language-specificity of a core component of Conceptual Structure. This in turn means for language evolution that Jackendoff's scenario, according to which Conceptual Structure pre-dates language and is shared among all higher animals and language has evolved as a system of external representations for Conceptual Structure, is now confronted with an increasingly plausible-looking alternative according to which Conceptual Structure in fact co-evolved with language as an interface between language and Spatial Structure.

Ultimately this for me is yet another take on the Whorfian hypothesis, the question to what extent language influences thought. Maybe we do think in a Conceptual Structure system, a symbolic system of cognitive reasoning that doesn't strictly depend on language but has in fact phylogenetically co-evolved with language. In that sense we have it thanks to language. Of course that also fits much more with a sort of Vygotskyian perspective of the relation between linguistic and cognitive development ontogenetically in children, according to which nonlinguistic thought is initially, in early stages, strongly influenced by linguistic development.

About the Series Editor

Fuyin (Thomas) Li (1963, Ph.D. 2002) received his Ph.D. in English Linguistics and Applied Linguistics from the Chinese University of Hong Kong. He is professor of linguistics at Beihang University, where he organizes *China International Forum on Cognitive Linguistics* since 2004, http://cifcl.buaa.edu.cn/Intro.htm. As the founding editor of the journal *Cognitive Semantics*, brill.com/cose, the founding editor of *International Journal of Cognitive Linguistics*, editor of the series *Distinguished Lectures in Cognitive Linguistics*, brill.com/dlcl, (originally *Eminent Linguists' Lecture Series*), editor of *Compendium of Cognitive Linguistics Research*, and organizer of ICLC-11, he plays an active role in the international expansion of Cognitive Linguistics.

His main research interests involve the Talmyan cognitive semantics, overlapping systems model, event grammar, causality, etc., with a focus on synchronic and diachronic perspective on Chinese data, and a strong commitment to usage-based model and corpus method.

His representative publications include the following: *Metaphor, Image, and Image Schemas in Second Language Pedagogy* (2009), *Semantics: A Course Book* (1999), *An Introduction to Cognitive Linguistics* (in Chinese, 2008), *Semantics: An Introduction* (in Chinese, 2007), *Toward a Cognitive Semantics, Volume I: Concept Structuring Systems* (Chinese version, 2017), *Toward a Cognitive Semantics, Volume II: Typology and Process in Concept Structuring* (Chinese version, 2019).

His personal homepage: http://shi.buaa.edu.cn/thomasli
E-mail: thomasli@buaa.edu.cn; thomaslfy@gmail.com

Websites for Cognitive Linguistics and CIFCL Speakers

All the websites were checked for validity on 20 January 2019.

Part 1 Websites for Cognitive Linguistics

1. http://www.cogling.org/
   Website for the International Cognitive Linguistics Association, ICLA

2. http://www.cognitivelinguistics.org/en/journal
   Website for the journal edited by ICLA, *Cognitive Linguistics*

3. http://cifcl.buaa.edu.cn/
   Website for China International Forum on Cognitive Linguistics (CIFCL)

4. http://cosebrill.edmgr.com/
   Website for the journal *Cognitive Semantics* (*ISSN 2352–6408/ E-ISSN 2352–6416*), edited by CIFCL

5. http://www.degruyter.com/view/serial/16078?rskey=fw6Q2O&result=1&q=CLR
   Website for the Cognitive Linguistics Research [CLR]

6. http://www.degruyter.com/view/serial/20568?rskey=dddL3r&result=1&q=ACL
   Website for Application of Cognitive Linguistics [ACL]

7. http://www.benjamins.com/#catalog/books/clscc/main
   Website for book series in Cognitive Linguistics by Benjamins

8. http://www.brill.com/cn/products/series/distinguished-lectures-cognitive-linguistics
   Website for Distinguished Lectures in Cognitive Linguistics (DLCL)

9. http://refworks.reference-global.com/
   Website for online resources for Cognitive Linguistics Bibliography

10. http://benjamins.com/online/met/
    Website for Bibliography of Metaphor and Metonymy

11. http://linguistics.berkeley.edu/research/cognitive/
Website for the Cognitive Linguistics Program at UC Berkeley

12. https://framenet.icsi.berkeley.edu/fndrupal/
Website for Framenet

13. http://www.mpi.nl/
Website for the Max Planck Institute for Psycholinguistics

Part 2 Websites for CIFCL Speakers and Their Research

14. CIFCL Organizer
Thomas Li, thomasli@buaa.edu.cn; thomaslfy@gmail.com
Personal homepage: http://shi.buaa.edu.cn/thomasli
http://shi.buaa.edu.cn/lifuyin/en/index.htm

15. CIFCL 18, 2018
Arie Verghagen, A.Verhagen@hum.leidenuniv.nl
http://www.arieverhagen.nl/

16. CIFCL 17, 2017
Jeffrey M. Zacks, jzacks@wustl.edu
Lab: dcl.wustl.edu
Personal site: https://dcl.wustl.edu/affiliates/jeff-zacks/

17. CIFCL 16, 2016
Cliff Goddard, c.goddard@griffith.edu.au
https://www.griffith.edu.au/griffith-centre-social-cultural-research/our-centre/cliff-goddard

18. CIFCL 15, 2016
Nikolas Gisborne, n.gisborne@ed.ac.uk

19. CIFCL 14, 2014
Phillip Wolff, pwolff@emory.edu

20. CIFCL 13, 2013 (CIFCL 03, 2006)
Ronald W. Langacker, rlangacker@ucsd.edu
http://idiom.ucsd.edu/~rwl/

21. CIFCL 12, 2013 (CIFCL 18, 2018)
Stefan Th. Gries, stgries@linguistics.ucsb.edu
http://www.stgries.info

22. CIFCL 12, 2013
Alan Cienki, a.cienki@vu.nl
https://research.vu.nl/en/persons/alan-cienki

23. CIFCL 11, 2012
Sherman Wilcox, wilcox@unm.edu
http://www.unm.edu/~wilcox

24. CIFCL 10, 2012
Jürgen Bohnemeyer, jb77@buffalo.edu
Personal homepage: http://www.acsu.buffalo.edu/~jb77/
The blog of the UB Semantic Typology Lab: https://ubstlab.wordpress.com/

25. CIFCL 09, 2011
Laura A. Janda, laura.janda@uit.no
http://ansatte.uit.no/laura.janda/

26. CIFCL 09, 2011
Ewa Dabrowska, ewa.dabrowska@northumbria.ac.uk

27. CIFCL 08, 2010
William Croft, wcroft@unm.edu
http://www.unm.edu/~wcroft

28. CIFCL 08, 2010
Zoltán Kövecses, kovecses.zoltan@btk.elte.hu

29. CIFCL 08, 2010
(Melissa Bowerman: 1942–2014)

30. CIFCL 07, 2009
Dirk Geeraerts, dirk.geeraerts@arts.kuleuven.be
http://wwwling.arts.kuleuven.be/qlvl/dirkg.htm

31. CIFCL 07, 2009
Mark Turner, mark.turner@case.edu

32. CIFCL 06, 2008
Chris Sinha, chris.sinha@ling.lu.se

33. CIFCL 05, 2008
Gilles Fauconnier, faucon@cogsci.ucsd.edu

34. CIFCL 04, 2007
Leonard Talmy, talmy@buffalo.edu
https://www.acsu.buffalo.edu/~talmy/talmy.html

35. CIFCL 03, 2006 (CIFCL 13, 2013)
Ronald W. Langacker, rlangacker@ucsd.edu
http://idiom.ucsd.edu/~rwl/

36. CIFCL 02, 2005
John Taylor, john.taylor65@xtra.co.nz
https://independent.academia.edu/JohnRTaylor

37. CIFCL 01, 2004
George Lakoff, lakoff@berkeley.edu
http://georgelakoff.com/